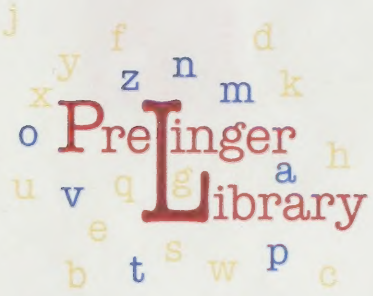
LIBRARY COPY

3781
MANNED SPACECRAFT CENTER
HOUSTON, TEXAS

SURPLUS - 1
LIBRARY OF CONGRESS
DUPLICATE

From the collection of the



San Francisco, California
2008

NCSA
16 m

THE BELL SYSTEM TECHNICAL JOURNAL

A JOURNAL DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

EDITORIAL BOARD

F. B. JEWETT
A. F. DIXON
D. LEVINGER
H. P. CHARLESWORTH
O. E. BUCKLEY
M. J. KELLY
W. WILSON
W. H. HARRISON
O. B. BLACKWELL
H. S. OSBORNE

R. W. KING, *Editor*
J. O. PERRINE, *Associate Editor*

TABLE OF CONTENTS AND INDEX

VOLUME XVII

1938

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

3781

PRINTED IN U. S. A.

THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XVII, 1938

Table of Contents

JANUARY, 1938

New Transmission Measuring Systems for Telephone Circuit Maintenance—*F. H. Best* 1
The Impedance Concept and its Application to Problems of Reflection, Refraction, Shielding and Power Absorption—*S. A. Schelkunoff* 17
On the Theory of Space Charge Between Parallel Plane Electrodes—*C. E. Fay, A. L. Samuel and W. Shockley* 49
A Carrier Telephone System for Toll Cables—*C. W. Green and E. I. Green* 80
Cable Carrier Telephone Terminals—*R. W. Chesnut, L. M. Ilgenfritz and A. Kenner* 106
Crystal Channel Filters for the Cable Carrier System—*C. E. Lane* 125
Crosstalk and Noise Features of Cable Carrier Telephone System—*M. A. Weaver, R. S. Tucker and P. S. Darnell* 137
A New Single Channel Carrier Telephone System—*H. J. Fisher, M. L. Almquist and R. H. Mills* 162

APRIL, 1938

Studies of Telephone Line Wire Spacing Problems—*J. A. Carr and F. V. Haskell* 195
Variable Equalizers—*H. W. Bode* 229
An Explanation of the Common Battery Anti-sidetone Subscriber Set—*C. O. Gibbon* 245
The Occurrence and Effect of Lockout Occasioned by Two Echo Suppressors—*Arthur W. Horton, Jr.* 258
Characteristic Time Intervals in Telephonic Conversation—*A. C. Norwine and O. J. Murphy* 281
Radioactivity—Artificial and Natural—*Karl K. Darrow* 292

JULY, 1938

Hertz, the Discover of Electric Waves—*Julian Blanchard* 327
Instruments for the New Telephone Sets—*W. C. Jones* 338
Transmission Features of the New Telephone Sets—*A. H. Inglis* 358
Spectrochemical Analysis in Communication Research—
Beverly L. Clarke and A. E. Ruehle 381
High Speed Motion Picture Photography—*W. Herriott* 393
An Optical Harmonic Analyzer—*H. C. Montgomery* 406
Magnetic Shielding of Transformers at Audio Frequencies—
W. G. Gustafson 416
Coaxial Cable System for Television Transmission—*M. E. Strieby* 438
Stabilized Feedback Oscillators—*G. H. Stevenson* 458
The Discovery of Electron Waves—*C. J. Davisson* 475

OCTOBER, 1938

Ultra-Short-Wave Transmission and Atmospheric Irregularities—
C. R. Englund, A. B. Crawford and W. W. Mumford 489
Amplitude Range Control—*S. B. Wright* 520
Devices for Controlling Amplitude Characteristics of Telephonic Signals—*A. C. Norwine* 539
The Exponential Transmission Line—*Chas. R. Burrows* 555
The Bridge Stabilized Oscillator—*L. A. Meacham* 574
Effect of Space Charge and Transit Time on the Shot Noise in Diodes—*A. J. Rack* 592
Fundamentals of Teletypewriters Used in the Bell System—
E. F. Watson 620
The Dielectric Properties of Insulating Materials—
E. J. Murphy and S. O. Morgan 640

The Bell System Technical Journal

Vol. XVII *January, 1938* *No. 1*

New Transmission Measuring Systems for Telephone Circuit Maintenance

By F. H. BEST

Transmission measurements on telephone circuits have long been recognized as essential aids in the furnishing of good service. Recent development work has produced new testing methods and apparatus which greatly simplify and expedite transmission measuring. This paper gives a brief description of the existing and the new arrangements.

ABOUT two decades ago pioneering work was being done in the development and introduction of methods and apparatus for the measurement of the transmission characteristics of local and toll telephone circuits, it having been realized that with the increasing complexity of the telephone plant, these measurements were necessary to insure satisfactory transmission performance. Transmission measurements now have a well established place in maintenance activity and development work is constantly in progress to improve the testing apparatus so that it can be operated more rapidly and conveniently, will cost less or do more things, and also to improve testing methods.

The principle employed in making transmission measurements has not changed. A standard testing power is supplied to one end of a telephone circuit and the power received at the other end is measured. The ratio between these powers, expressed in decibels, is a measure of the transmission loss or gain in the circuit. The magnitude of the testing powers is small, the sending power for measuring being one-thousandth of one watt and the received power ranging from one ten-thousandth to one-millionth of a watt. Sensitive meters must be used, or the power amplified before measuring it. Direct-current meters are more sensitive than alternating-current meters and it is the practice to convert the weak received alternating current to direct current by means of vacuum tube or copper-oxide rectifiers and employ direct-current meters for measuring it.

Until recently the developments have been along the line of improvements in the apparatus itself, there being practically no change

in the general methods and arrangements employed. For the large toll offices where much testing is done the measuring apparatus has been installed at one or more points in an office, the installations being known as transmission test boards. From these points trunks radiate to repeaters, test boards, switchboards, etc., and the circuits and equipment to be tested are connected to them by patching cords or switchboard cords. The testing has been done by a trained force of transmission testers, this force often being separate from that of the test board attendants whose work has consisted chiefly in correcting line faults, signaling troubles, etc., which do not require transmission measurements for their location.

In the local plant, testing has been done almost entirely with portable transmission measuring apparatus in the hands of a transmission testing force who travel from office to office, testing the cord circuits, switching trunks, etc., at intervals of one or two years. The measuring apparatus has been costly and its permanent installation in each office could not be justified. The apparatus used in both local and toll testing has been described in a number of articles.[1, 2, 3]

During the last few years new methods and apparatus have been developed and radical changes made that greatly improve the situation both from an economic and operating standpoint. The new measuring instruments are much less costly than the types heretofore available and are as stable in operation and as simple to use as ammeters and voltmeters. In the local plant the decrease in cost and complexity and improvement in operating methods has justified permanent installations in many of the larger offices while the newer forms of portable apparatus are being much more extensively distributed than earlier types. For the toll plant the improvement consists in doing away largely with the centralized transmission testing point. The regular test board force now can make transmission measurements at their test boards and the repeater attendants can measure the performance of the repeaters while working on them. The work is so simple and the maintenance forces are so well trained that a special transmission testing force is not required.

This change has been brought about by the development of new instrumentalities and improved circuits, chief among which are more sensitive meters, the copper-oxide rectifier and the negative feedback amplifier.[5] Until recently, the most sensitive meters which were suitable for general use would not measure the weak power used in transmission measuring so that amplification of this power was required. Using the alloy steels now available, meters of much greater sensitivity and equal ruggedness have been manufactured and transmission losses

of 20 db can be measured without an amplifier and without using abnormal testing power. Vacuum tubes formerly used to rectify the received alternating current testing power have been largely superseded by the copper-oxide rectifier which is small, inert and requires no added power for its operation. The combination of this rectifier with sensitive meters greatly simplifies and reduces the cost of the measuring apparatus. Figure 1 shows the simplicity of the circuit of a transmission measuring set which will measure losses up to 20 db.

While indicating meters have been used in transmission measuring for many years, it is only recently that they have been calibrated directly in db. This is the preferable way of measuring and its adoption has been delayed only because of the limitations of available apparatus and circuits which prevented a stable device from being developed. The earlier copper-oxide rectifiers were too unstable for

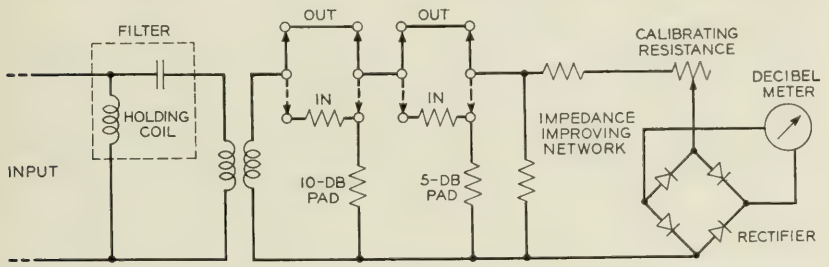


Fig. 1—Simplified circuit of transmission measuring set having a range of 20 db without an amplifier.

measuring and until the development of the negative feedback amplifier, all vacuum tube amplifiers varied in amplification with changes in filament and plate voltage and the aging of tubes. The voltage of available power plants, while sufficiently stable for commercial use, caused the pointer of the meter in a transmission measuring set to fluctuate and over intervals of a few hours the change might be as much as one db. Smaller changes occurred frequently. These changes were slow enough so that measurements could be made, but they required frequent adjustment of amplifier gain to compensate. With such instability, the most rapid and accurate measurements can be made by using the comparison method of testing where a known "standard" power is attenuated by calibrated potentiometers or networks until it equals the unknown received testing power. A meter in this case serves merely as a means for telling when the two are equal and the result is read from calibrated dials on the attenuator. A majority of the measuring sets now in the plant operate on this

"flip flop" principle. The use of these dials generally requires that the measuring instrument be built as a unit and that measurements be made at its location. The latter requirement often prevents its installation at the most desirable point because space is not available.

For toll circuit testing in the larger offices where a considerable number of routine and trouble locating tests are made, the required frequency range and loss range are beyond the scope of a simple copper-oxide rectifier and meter. The ability to make level measurements is also a desirable feature. To meet these requirements, an amplifier is provided in the receiving circuit. By applying the negative feedback principle to the amplifier and rectifier of the latest toll transmission measuring system a remarkably stable circuit has been obtained so that the meter can be calibrated directly in db and oper-

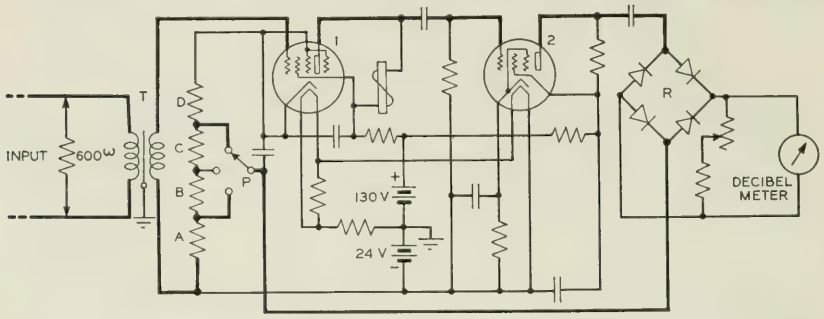


Fig. 2—Simplified circuit of reverse feedback amplifier rectifier used for toll circuit maintenance.

ated for long periods without adjustment. This circuit, which is shown in simplified form in Fig. 2, consists of a high-impedance input transformer T bridged across a 600-ohm terminating resistance which may be removed when level measurements are made, two pentode tubes 1 and 2, a copper-oxide rectifier R and a meter M. This combination has much more amplification than is required, so the excess is used to improve stability by introducing part of the output voltage into the input circuit of the first tube in such phase relation with respect to the applied input voltage that the net input voltage is reduced. This reverse or negative feedback voltage is introduced into the grid circuit of the first tube through resistances A, B and C by connecting one of the rectifier terminals to the movable arm of the potentiometer P. These resistances, and resistance D form the cathode drop resistance of the grid circuit and any potential applied across them affects the potential on the grid of the tube. In the position shown, the reverse feedback is a maximum and the net amplification

of the amplifier is a minimum. Moving the potentiometer arm to the lowest step gives less feedback and the amplifier net gain is greater.

The principle of stabilization is as follows: If a constant potential is applied to the input terminals of the amplifier a constant potential will also be applied to the grid of the first tube. As long as the tubes or the rectifier do not change in characteristics the output of the rectifier will likewise be constant. Now if through any cause, such as a change in tube characteristics or in the rectifier, the gain of the amplifier-rectifier should increase, the output will increase and the neutralizing voltage fed back into the input circuit will also increase. This will reduce the voltage on the grid so that the net output voltage of the amplifier-rectifier will not be changed noticeably. Conversely, if a change in the tubes or the rectifier should cause a reduction in gain, the voltage fed back into the input circuit will decrease, there will be less feedback voltage and the output voltage will be substantially the same as before. The reverse feedback amplifier-rectifier is so stable that once it is adjusted to have the proper characteristics it will remain constant for long periods.

The meters used with this amplifier-rectifier have a range of 15 db. This is less than the required measuring range so that it is necessary to increase it by changing the amplifier gain in steps of 10 db. The reverse feedback amplifier lends itself readily to this as the variation of a single resistance in the feedback circuit is sufficient and no expensive balanced attenuators are required. In Fig. 2 this is done by potentiometer *P*. In practice this resistance is controlled by relays which form a part of the amplifier, these relays in turn being remotely controlled by keys, jacks or dials at various points in the office.

The db meters can be located where desired without reference to the location of the amplifier-rectifier. They may also be placed in lantern slide projectors which throw a greatly enlarged meter scale on a screen so that it can be read from distances up to 50 feet or more. The new arrangements are extremely flexible, one set of equipment supplying measuring facilities for several different points in an office where previously several sets would be necessary. Where the use of the equipment is intermittent more than one meter may be used with a single amplifier-rectifier. The meters from which the results are read may be of the conventional indicating type which can be mounted on keyshelfs or on vertical panels, they may be of the projector type or they may be of the recording type which records on paper the characteristics which are being measured. All meters are interchangeable, being similar electrically.

These new instrumentalities have removed many of the limitations

to the design of measuring systems and the recently developed arrangements are therefore better coordinated with other facilities than those which they replace. A few examples will be given to illustrate the recent advances in transmission measuring work.

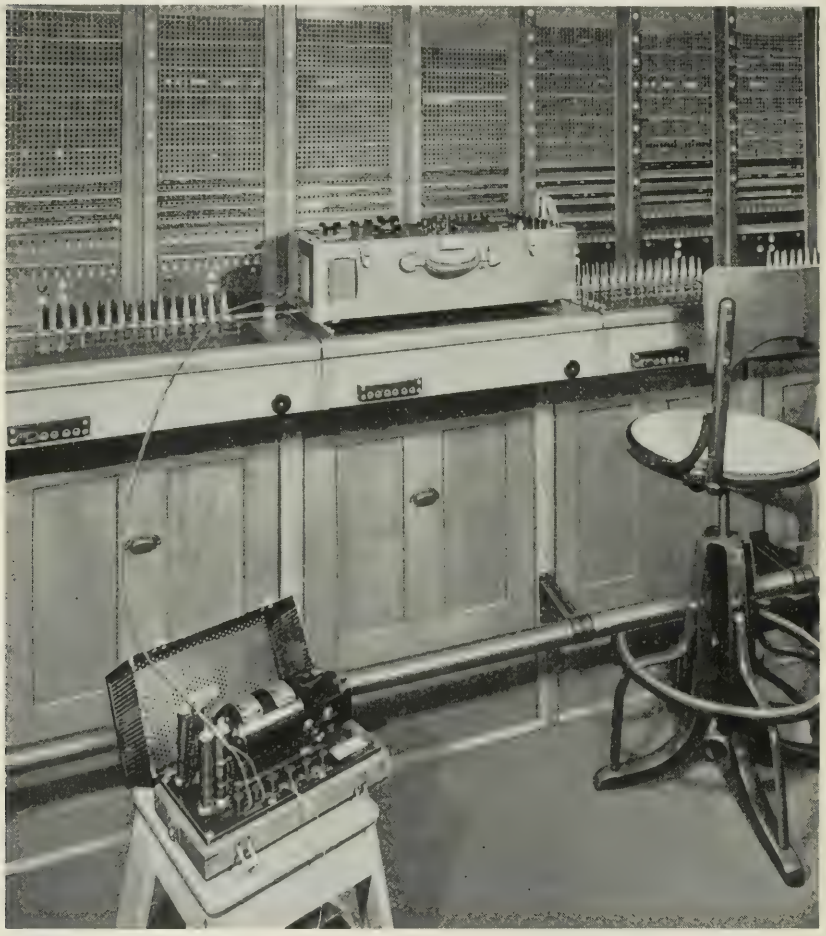


Fig. 3—Transmission measuring set developed about 1920 and associated generator as used at a switchboard in testing cord circuits or interoffice trunks. Both units are required for either sending or receiving.

Figure 3 shows the testing power generator and measuring set developed about 1920 for use in the local plant to measure central office equipment and interoffice trunks with testing power of a single frequency, the frequency employed usually being 1000 cycles. When

measuring circuits between offices, it is necessary either to have duplicate sets of equipment at the two ends of the circuit or to connect two circuits together at one end, testing them at the other end as one circuit. This latter arrangement, while not wholly satisfactory, has been the one generally employed because of the greater expense of the

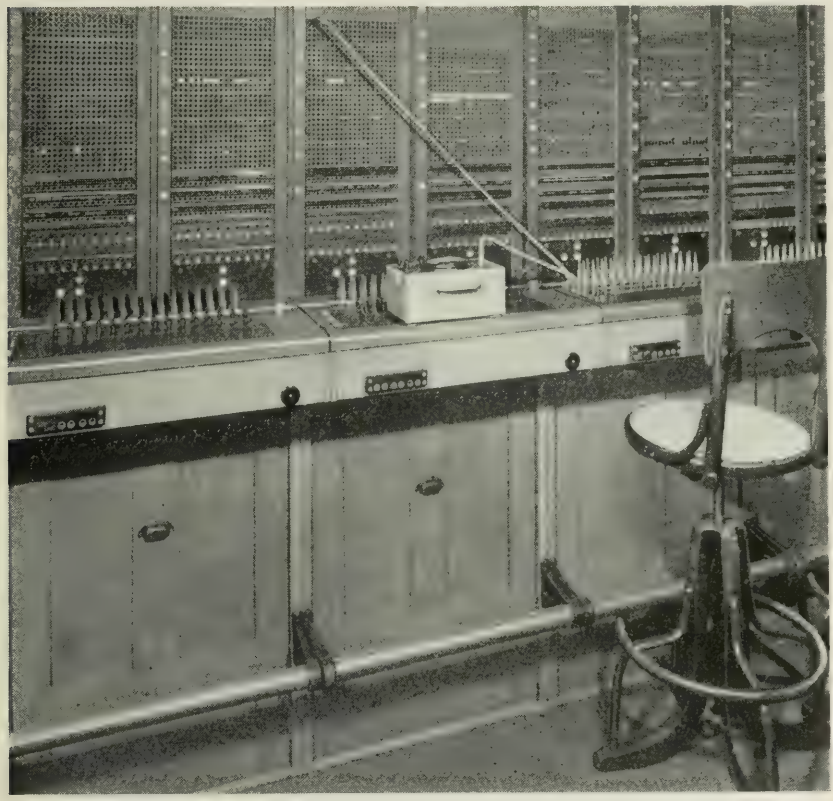


Fig. 4—Latest type of transmission measuring set as used at a switchboard. The associated generator shown in Fig. 5 is usually permanently mounted in the office and testing power is obtained through the multiple as shown. For one-way measurements on circuits between offices, the generator is connected to one end of the circuit and the transmission measuring set at the other, both units not being required at each end.

other method which has involved not only two sets of expensive testing apparatus but two testers, these being necessary because the testing apparatus required frequent adjustment. If circuits between a group of offices in a city are to be tested between circuit terminals rather than by the looping back method, considerable time is consumed in traveling between offices when using this apparatus.

The new apparatus shown in Figs. 4 and 5 not only costs but one-tenth as much and weighs one-quarter as much as the older apparatus but also has electrical advantages which enable a new and improved measuring technique to be employed. The generator is a magneto inductor alternator driven by a 50-cycle or 60-cycle induction motor and gives a constant output without attention. The output is adjusted at the factory or on installation. Some of these machines have

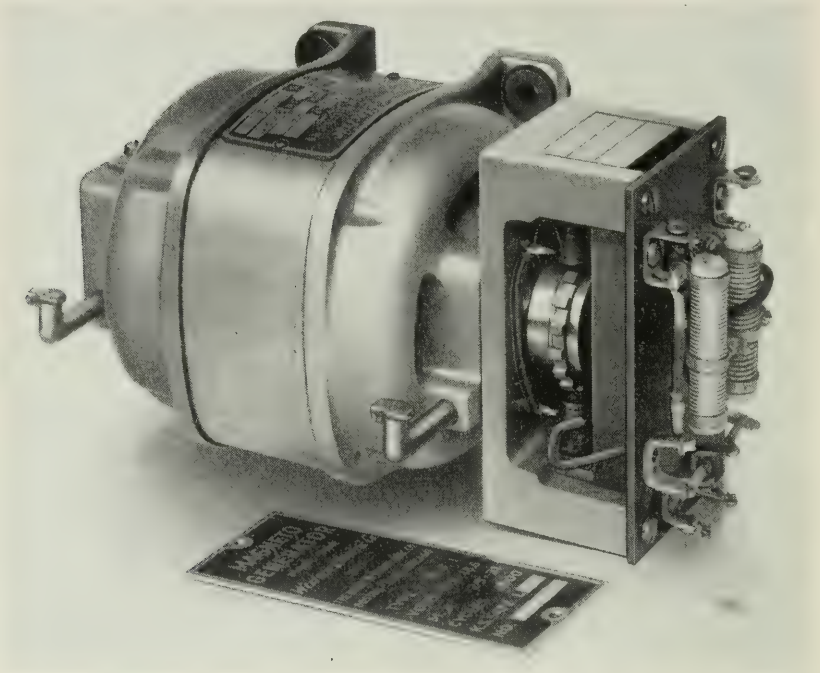


Fig. 5—One thousand-cycle magneto generator used generally for transmission testing. Cover plate removed to show the generator construction. Overall length 7 inches.

been running continuously for about a year without showing any appreciable change in output. They can, therefore, be mounted permanently in an office and the output terminals wired to convenient testing points so that the generator need not be carried around. Because of this output stability, it is not only unnecessary to have an experienced tester at the distant end of the circuit but in the larger offices auxiliary switching equipment is arranged so that testing power can be supplied automatically to one end of the circuit by direction of the tester at the other end, who simply calls or dials a designated

number over the circuit to be tested.[7] The testing power is cut off at the sending end when the connection is broken at the receiving end. Tests can be made in the same manner from private branch exchanges and subscribers' stations without a charge being registered against the subscriber.

Fig. 6—Small receiving set as used to measure the transmission loss of a subscriber's line. The 1000-cycle generator has been connected to the central office end of the line by calling a designated number. When the telephone instrument has been replaced on the mounting, the meter in the measuring set will read the line loss.

The receiving set shown in Fig. 4 is based on the electrical circuit of Fig. 1. It has a transmission range of 20 db, and is provided with all the jacks and facilities required for testing cord circuits, trunks and central office equipment. A still smaller and less expensive receiving set of a similar type having a 10 db range is shown in Fig. 6. It is extremely portable and light and is useful for work where the limited

range is sufficient. Jacks and other testing conveniences have been omitted to save bulk and cost.

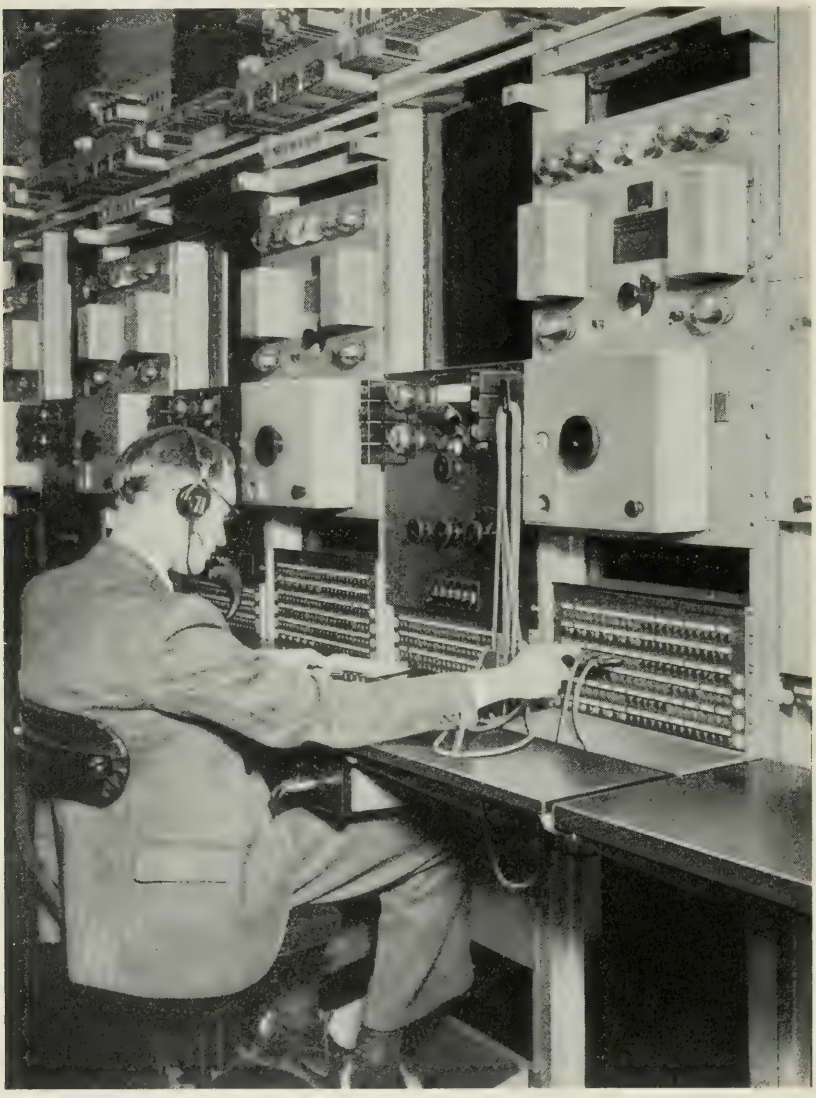


Fig. 7—Transmission testboard for maintaining long distance circuits.

The new instruments for the local plant are not limited to local testing but also have extensive application in the toll plant especially in the smaller offices which do not have many toll circuits. Because

of its excellent performance the 1000-cycle generator shown in Fig. 5 will be used generally in the toll plant for 1000-cycle testing.

Figure 7 shows several positions of the transmission test board now widely used for measuring toll telephone circuits. Introduced in 1926, it preceded the development of the reverse feedback amplifier-rectifier and the transmission measuring set used in this board is therefore of the comparison type in which the results are read fron calibrated dials. Testing power is provided by a variable frequency oscillator from which frequencies in the voice range can be obtained. The size of this equipment precluded its being installed in test boards which contain the circuit terminals so that trunks between the measuring

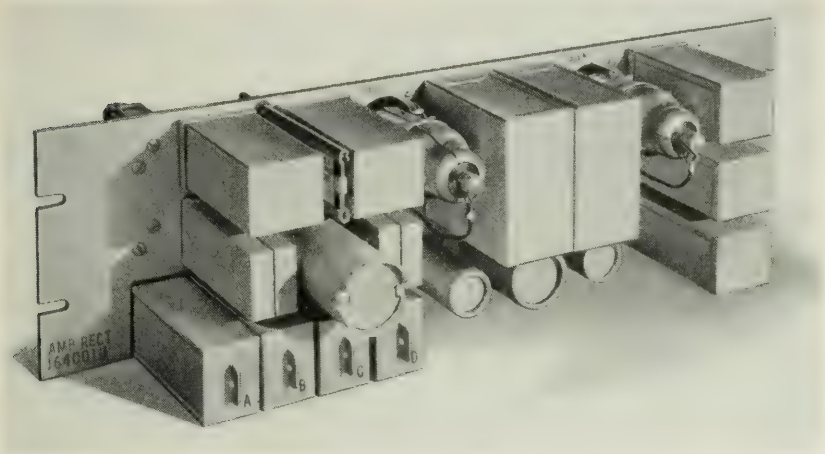


Fig. 8—New amplifier rectifier for general use in toll transmission maintenance.

equipment and these terminals are necessary. In large offices a number of these transmission test boards have been provided.

Figure 8 shows the new reverse feedback amplifier-rectifier used with the latest toll transmission measuring system. This simple panel, which is about one-fourth the size of the measuring set employed in the test board shown in Fig. 7, contains everything required at the receiving end of a circuit excepting the meter and the keys for changing the measuring range. With its associated meter and keys, it is less than one-half as expensive as corresponding elements in the transmission testboards. It can be used with the variable frequency oscillator shown in Fig. 7 or with the 1000-cycle machine shown in Fig. 5.

The meters used with this amplifier-rectifier have a specially designed magnetic circuit which gives an evenly spaced scale on the

meter.[8] A conventional meter would have large db divisions at one end of the scale and small ones at the other.

When a direct reading transmission measuring set is used to make measurements over a wide frequency range, it is essential that the response be the same at all frequencies. Variation of response of the new feedback amplifier-rectifier with frequency is so small that it can be calibrated with 1000-cycle current and measurements made over a wide frequency range without recalibration.

From the standpoint of efficiency, the best place for making trans-

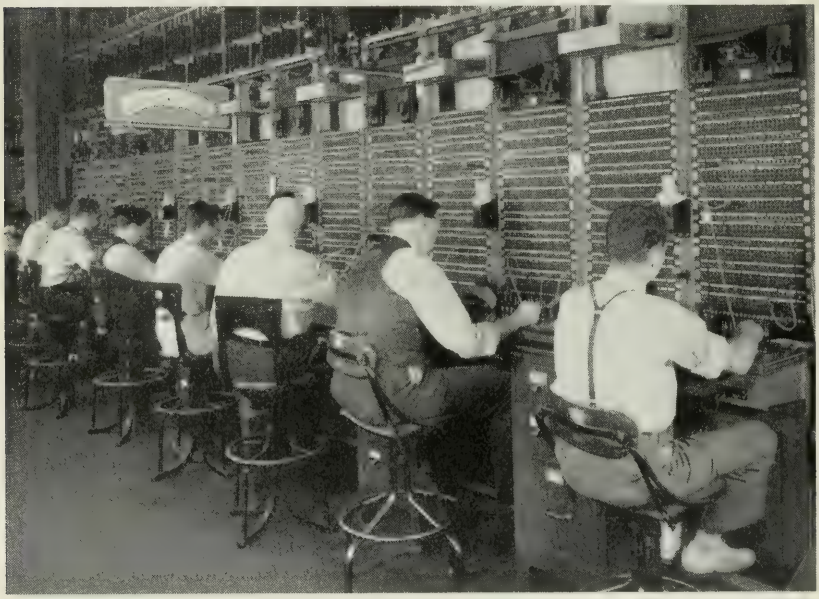


Fig. 9—Projection meter as used for measuring transmission at a secondary testboard. The meter is being read by the third man from the right.

mission measurements on complete toll circuits is at the test board to which all troubles are reported by the operator and where overall signaling and other operating tests are made. Practically all of the space in this board is taken up by jacks on which the circuits terminate so that the older types of measuring set could not be provided at this point. The new arrangement is ideal for application to this type of board since all that is required in the board are the meter and a few jacks or keys for controlling its range. This type of measuring device can be applied equally well to new and existing boards of various types. This feature is of particular value as the maximum efficiency of high-speed measuring systems can be obtained only when all test boards

are equipped with them. Two arrangements are available for test board use. One employs a conventional type of meter mounted in the keyshelf or on a panel and the other employs the projection type of meter which is illustrated in Fig. 9.[6] The method of operation is simple. When a transmission test is to be made the tester listens until he hears the tone caused by testing power coming over the circuit from the distant generator, then connects the circuit to a jack in which the measuring set input terminates. The meter indicates immediately the net loss of the circuit at the testing frequency. With the projection meter arrangement the lamp in the projector is turned on automatically when a connection is made to the test jack.

The new amplifier-rectifier will measure transmission losses and

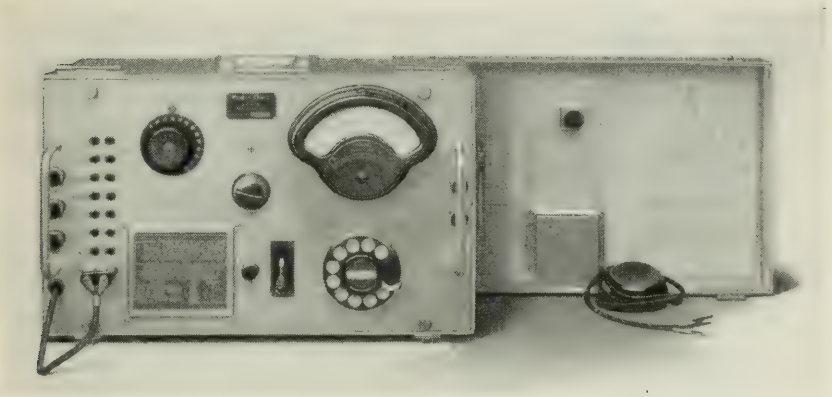


Fig. 10—Noise measuring set.

gains and also transmission level. For the latter type of measurement the input impedance of the amplifier is raised to a high value so that it is, in effect, a voltmeter. This change in impedance can be made from a remote point, a relay for making the change being a part of the amplifier.

In addition to the measurement of transmission losses, gains and levels on telephone circuits, it is also necessary at times to make measurements of noise on the circuit. This noise may be caused by currents induced by power systems or from power plants in the telephone offices or it may be in the form of crosstalk, sometimes unintelligible, from other telephone circuits. Noise measurements are now made with meter indicating devices, the latest type of self-contained portable noise measuring set being shown in Fig. 10.[9] Where enough noise measurements are made to justify a permanent installation, an arrangement similar to that described for transmission

measurements can be used. While a different amplifier-rectifier is required for noise measurements the same meters and methods of control can be employed as for transmission measurements.

All of the methods so far described are manually operated in so far as the recording of the results is concerned. There are occasions when a fully automatic recording device is desirable.[4] Such cases are the making of transmission versus frequency runs on repeaters or circuits where measurements are desired over a wide range of frequencies. Another class of measurements are those in which the single-frequency transmission loss of a circuit is to be determined over a long period of time to obtain a measure of the stability of the circuit. For this purpose the new method of measuring transmission is well adapted. With a stable amplifier-rectifier having practically no frequency distortion, the indicating meter may be replaced by a recording meter which will record continuously the received power expressed in db. A fully automatic recording system is shown in Fig. 11. With this arrangement an oscillator at one end of the line supplies testing power to the line, the frequency of the power being changed continuously by a synchronous motor. At the receiving end the recording meter, also operated by a synchronous motor, plots the received power. A complete transmission frequency run on a message telephone circuit can be made in less than one minute. When records are to be made of transmission vs. time, the oscillator frequency at the sending end is fixed and the meter plots the transmission loss at that frequency.

The above description has been limited to the types of measurements commonly made on complete circuits or parts of circuits. In addition to these there are a number of types of measuring apparatus which are used in connection with the installation of new equipment, changes in installations and the detailed running down of trouble.

Transmission measurements have proved to be of great value in maintaining satisfactory transmission performance of telephone circuits. Periodic routine tests avoid service impairment by detecting troubles. Troubles which are thus detected or which cause service complaint are located readily. Another large field of use is in the adjustment of repeaters and complete circuits to prescribed transmission characteristics. The importance of this work to the Bell System is indicated by the fact that there are nearly 1000 of the portable transmission measuring equipments and 1500 transmission test boards now in use, with which several million measurements are made annually. The new measuring systems enable this work to be done more rapidly than in the past, and the reduced cost of the equipment is resulting in its greater distribution.

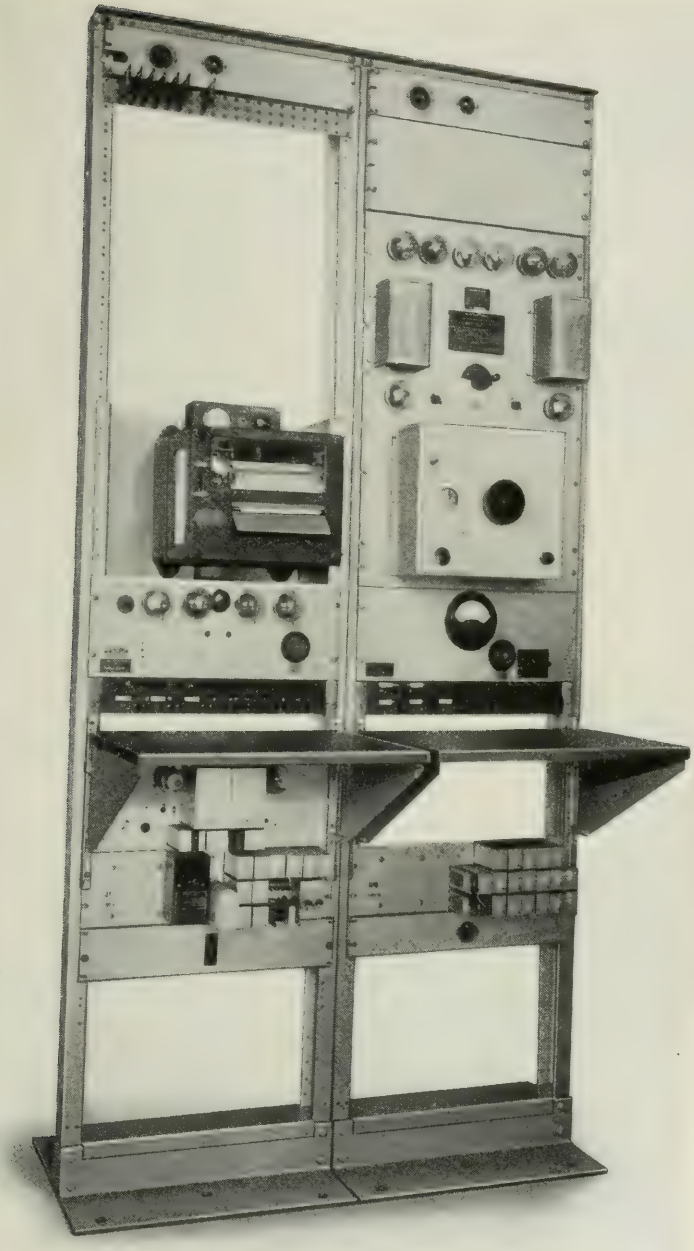


Fig. 11—Automatic recording transmission measuring system.

References

1. "Measuring Methods for Maintaining the Transmission Efficiency of Telephone Circuits," F. H. Best, *Elec. Engg.*, vol. 43, February 1924, pp. 136–144.
2. "Electrical Tests and Their Applications in the Maintenance of Telephone Transmission," W. H. Harden, *Bell System Technical Journal*, vol. 3, July 1924, pp. 353–392.
3. "Practices in Telephone Transmission Maintenance Work," W. H. Harden, *Elec. Engg.*, vol. 43, December 1924, pp. 1124–1128.
4. "A Recording Transmission Measuring System for Telephone Circuit Testing," F. H. Best, *Bell System Technical Journal*, vol. 12, January 1933, pp. 22–34.
5. "Stabilized Feedback Amplifiers," H. S. Black, *Bell System Technical Journal*, vol. 13, January 1934, pp. 1–18.
6. "Projecting Circuit Performance on a Screen," F. E. Fairchild, *Bell Laboratories Record*, vol. 13, July 1935, pp. 328–331.
7. "Improved Transmission Measuring System," F. H. Best, *Bell Laboratories Record*, vol. 14, March 1936, pp. 237–239.
8. "Decibel Meters," F. H. Best, *Bell Laboratories Record*, vol. 15, January 1937, pp. 167–169.
9. "A New Noise Meter," J. M. Barstow, *Bell Laboratories Record*, vol. 15, April 1937, pp. 252–256.

The Impedance Concept and Its Application to Problems of Reflection, Refraction, Shielding and Power Absorption

By S. A. SCHELKUNOFF

This paper calls attention to the practical value of a more extended use of the impedance concept. It brings out a certain underlying unity in what otherwise appear diverse physical phenomena. Although an attempt has been made to trace the history of the concept of "impedance" and many interesting early suggestions have been found, reference to these lies beyond the scope of this paper. Apparently, Sir Oliver Lodge was the first to use the word "impedance," but the concept has been developed gradually as circumstances demanded through the efforts of countless workers.

The main body of the paper is divided into three parts: Part I, dealing with the exposition of the impedance idea as applied to different types of physical phenomena; Part II, in which the general formulæ are deduced for reflection and transmission coefficients; Part III, presenting some special applications illustrating the practical utility of the foregoing manner of thought.

THE term "impedance" has had an interesting history, in which one generalization has suggested another with remarkable rapidity. Introduced by Oliver Lodge,[1] it meant the ratio V/I in the special circuit comprised of a resistance and an inductance, I and V being the *amplitudes* of an alternating current and the driving force which produced it. This was soon extended to the somewhat more general circuit consisting of a resistance, an inductance coil and a condenser.[2] The usage did not develop much further until the use of

[1] Dr. Oliver Lodge, F.R.S., "On Lightning, Lightning Conductors, and Lightning Protectors," *Electrical Review*, May 3, 1889, p. 518

[2] It is interesting to note that the first impulse was to introduce a new word rather than to extend the meaning of the old term. Thus in 1892, F. Bedell and A. Crehore write as follows: "From the analogy of this equation to Ohm's law, we see that the expression $\sqrt{R^2 + \left(\frac{1}{C\omega} - L\omega\right)^2}$ is of the nature of a resistance, and is the apparent resistance of a circuit containing resistance, self-inductance and capacity. This expression would quite properly be called 'impedance' but the term impedance has for several years been used as a name for the expression $\sqrt{R^2 + L^2\omega^2}$, which is the apparent resistance of a circuit containing resistance and self-inductance only. We would suggest, therefore, that the word 'impediment' be adopted as a name for the expression $\sqrt{R^2 + \left(\frac{1}{C\omega} - L\omega\right)^2}$ which is the apparent resistance of a circuit containing resistance, self-induction and capacity, and the term impedance be retained in the more limited meaning it has come to have, that is $\sqrt{R^2 + L^2\omega^2}$, the

complex quantities, which had begun early in the nineteenth century among mathematicians, was popularized among engineers by Kennelly and Steinmetz. Then the proportionality relation $V = ZI$, which had previously been true only if V and I were interpreted as amplitudes, acquired a more general significance, for it was found that this relation could express the phase relationship as well, provided Z was given a suitable complex value.

An important generalization came when the close similarity of the laws connecting V and I in an electric circuit to those governing force and velocity in mechanical systems suggested that the ratio "force/velocity" be called a "mechanical impedance." This usage is now well nigh universal.

The next step was a short one: it amounted to extending the term to include also the ratio "force per unit area/flow per unit area"; that is, "pressure/flux." This usage is well known in such fields as acoustics, but it has not penetrated as far into the electrical field as convenience seems to warrant.

If we read these remarks with a view to appraising the direction in which future growth might be expected, we are immediately impressed by the strong trend toward interpreting the ratio "force/velocity" in an ever widening sense. It is my purpose in the present paper to indicate some further extensions which I have found to be useful. They are founded upon five basic ideas. The first is to recognize and use whenever possible analogies between dynamical fields in which the impedance concept is common and others (heat, for instance) in which it is not. The second is the idea of extending the V/I relation from circuits to radiation fields, in much the same way that the "force/velocity" concept has been made to embrace "pressure/flux" in hydrodynamics. The third is, to regard the impedance as an attribute of the *field* as well as of the body or the medium which supports the field, so that the impedance to a plane wave is not the same as the impedance to a cylindrical wave, even when both are propagated in infinite "free space." The fourth basic idea is that of assigning *direction* to the impedances of fields. This does not mean, however, that the impedances are vectors; in fact, they are not, since they fail to obey the laws of addition and the laws of transformation peculiar to vectors. And finally the fifth is a generalization of the idea of a *one-dimensional transmission line* or simply a *transmission line.* While

apparent resistance of a circuit containing resistance and self-induction only." Frederick Bedell and Albert C. Crehore, "Derivation and Discussion of the General Solution of the Current Flowing in a Circuit Containing Resistance, Self-Induction and Capacity, With Any Impressed Electromotive Force," *Journal A. I. E. E.*, Vol. IX, 1892, pp. 303–374, see p. 340.

all physical phenomena are essentially three-dimensional, frequently all but one are irrelevant and can be ignored or are relatively unimportant and can be neglected. In the mathematical language, this means that only one coordinate (distance, angle, etc.) is retained explicitly in the *equations of transmission.*

The paper is divided into three parts. Part I discusses broadly the ratios to which the term "impedance" can appropriately be applied in a wide variety of physical fields, ranging from electric circuits and heat conduction to electromagnetic radiation. In this part the concept is gradually broadened until at the end it has acquired the property of direction mentioned above. Parts II and III consider the general laws governing reflection, refraction, shielding and power absorption, and rephrase them as theorems regarding the generalized impedances. To make the illustrations more effective, familiar examples are chosen.

PART I

THE IMPEDANCE CONCEPT

Electric Circuits

In an electric circuit comprised of a resistance R and an inductance L, the instantaneous voltage-current relation is described by the following differential equation

$$L\frac{dI_0}{dt} + RI_0 = V_0, \tag{1}$$

where V_0 is the applied electromotive force. If V_0 varies harmonically with frequency f, ultimately I_0 will also vary harmonically with frequency f. What happens is that the solution of (1) consists of two parts, the *transient* part and the *steady state* part, the former decreasing exponentially with time and the latter being periodic.

The steady state solution of (1), or indeed of the most general linear differential equation with constant coefficients, can be found by means of a simple mathematical device based upon the use of complex numbers. Thus if V_0 and I_0 vary harmonically, they may be regarded as real parts of the corresponding complex expressions $Ve^{i\omega t}$ and $Ie^{i\omega t}$, where $f = \omega/2\pi$ is the frequency. The quantities V and I are complex numbers whose moduli represent the amplitudes and whose phases are the initial phases (at the instant $t = 0$) of the electromotive force and the electric current. The time rate of change of I_0 is then the real part of the derivative of $Ie^{i\omega t}$, that is, the real part of $i\omega Ie^{i\omega t}$.

If we form another equation after the pattern of (1), replacing I_0 and V_0 by the *imaginary* part of $Ie^{i\omega t}$ and $Ve^{i\omega t}$, and add the new

equation to (1), we shall have

$$L\frac{d(Ie^{i\omega t})}{dt} + R(Ie^{i\omega t}) = V_0 e^{i\omega t}.$$

Differentiating and cancelling the time factor $e^{i\omega t}$, we obtain

$$(R + i\omega L)I = V.$$

The ratio $Z = V/I = Ve^{i\omega t}/Ie^{i\omega t}$ is called the *impedance* of the electric circuit. In the present instance

$$Z = R + i\omega L.$$

In general, the impedance $Z = R + iX$ has a real and an imaginary part, the former being the *resistive* component of the impedance and the latter the *reactive.*

Mechanical Circuits

Linear oscillations of a mass in a resisting medium are described by equations identical with (1) and (2) except for the customary difference in lettering

$$m\frac{d(ve^{i\omega t})}{dt} + r(ve^{i\omega t}) = Fe^{i\omega t}.$$

In this equation, v represents the velocity and F the applied force, m the mass and r the resistance coefficient. The mechanical impedance is then

$$Z = r + i\omega m.$$

Similarly, for torsional vibrations the impedance is defined as the ratio "torque/angular velocity."

Electric Waves in Transmission Lines

Let x be the distance coordinate specifying a typical section of an electric transmission line. Let the complex quantities V and I be the voltage across and the electric current in the transmission line.[3] Then the space rate of change of the voltage is proportional to the current and the space rate of change of the current is proportional to the voltage

$$\frac{dV}{dx} = -ZI, \qquad \frac{dI}{dx} = -YV. \tag{2}$$

[3] The time factor $e^{i\omega t}$ is usually implicit.

The coefficients of proportionality Z and Y are known as the distributed *series impedance* and the distributed *shunt admittance* of the line; they depend upon the distributed series resistance R, shunt conductance G, series inductance L and shunt capacity C in the following manner:

$$Z = R + i\omega L, \qquad Y = G + i\omega C. \tag{3}$$

In a generalized transmission line Z and Y may be functions of x and may depend upon ω in a more complicated manner than that suggested in (3).

If Z and Y are independent of x, (2) possesses two exponential solutions:

$$\begin{aligned} I^+ &= Ae^{-\Gamma x + i\omega t}, & V^+ &= Z_0 I^+; \\ I^- &= Be^{\Gamma x + i\omega t}, & V^- &= -Z_0 I^-; \end{aligned} \tag{4}$$

where

$$\Gamma = \alpha + i\beta = \sqrt{ZY}, \qquad Z_0 = \sqrt{\frac{Z}{Y}} = \frac{\Gamma}{Y} = \frac{Z}{\Gamma}.$$

It is customary to designate by Γ that value of the square root which is in the first quadrant of the complex plane or on its boundaries; the other value of the square root is $-\Gamma$.

The two "secondary" constants Γ and Z_0 are called, respectively, the *propagation constant* and the *characteristic impedance.* The real part α of the propagation constant is the *attenuation constant* and β is the *phase constant.*

Equations (4) represent *progressive* waves because an observer moving along the line with a certain finite velocity beholds an unchanging phase of V and I. This velocity c is called the *phase velocity* of the wave. Setting $x = ct$ in the upper pair of (4), we obtain the condition for the stationary phase

$$-\beta c + \omega = 0, \qquad c = \frac{\omega}{\beta}.$$

Hence, V^+ and I^+ represent a wave traveling in the positive x-direction. Similarly we find that V^- and I^- represent a wave traveling in the opposite direction.

Consider two points in which the phases of V and I differ by 2π when observed at the same instant; the distance λ between these points is called the wave-length. By definition

$$\beta\lambda = 2\pi, \qquad \lambda = \frac{2\pi}{\beta}.$$

If the transmission line is non-uniform, that is, if Z and Y are functions of x, then the solutions of (2) are usually more complicated. In any case, however, there are two linearly independent solutions $I^+(x)$ and $I^-(x)$ in terms of which the most general solution can always be expressed

$$I(x) = AI^+(x) + BI^-(x).$$

These independent solutions may represent either progressive waves in two opposite directions or certain convenient combinations of such waves.

The corresponding V-functions are found by differentiation from (2); thus

$$V^+(x) = -\frac{1}{Y}\frac{dI^+}{dx}, \qquad V^-(x) = -\frac{1}{Y}\frac{dI^-}{dx}.$$

The *impedance* of the V^+, I^+-wave is then

$$Z_0^+(x) = \frac{V^+(x)}{I^+(x)} = -\frac{1}{YI^+}\frac{dI^+}{dx} = -\frac{1}{Y}\frac{d}{dx}(\log I^+).$$

Similarly the impedance of the V^-, I^--wave is

$$Z_0^-(x) = -\frac{V^-(x)}{I^-(x)} = \frac{1}{YI^-}\frac{dI^-}{dx} = \frac{1}{Y}\frac{d}{dx}(\log I^-). \tag{5}$$

The negative sign in (5) is merely a matter of convention: the "positive" and the "negative" directions of the transmission line are so defined that the real parts of Z_0^+ and Z_0^- are positive.

In general, Z_0^+ and Z_0^- are not equal to each other. Moreover, there is a considerable amount of arbitrariness in our choice of the basic solutions I^+ and I^-. Thus, we are brought face to face with the fact that we must regard the impedance as an attribute of the wave as well as of the transmission line. This point of view will become even more prominent when we come to deal with the wave transmission in three-dimensional media. There even progressive waves may have different characters (they may be plane, cylindrical, spherical, etc.) and the impedances of the *same* medium to these waves will be different. And naturally, it goes without saying that the impedances to like waves in *different* media may also be different. One could, perhaps, take the position that geometrically similar waves in different media are not really alike if the corresponding "force/velocity" ratios are not equal and that under all circumstances the "impedance" is the property of a wave. However, "intrinsic im-

pedance" will be used to designate a constant of the medium without reference to any particular wave.

Vibrating Strings

In strings under constant tension τ, simply periodic waves may be described by the following two equations:

$$\frac{dF}{dx} = -(r + i\omega m)v, \qquad \frac{dv}{dx} = -\frac{i\omega}{\tau}F,$$

where m is the mass and r the resistance per unit length of the string. The variable F represents the force on a typical point of the string at right angles to the string and v is the velocity at that point.

Hence the characteristic impedance and the propagation constant are given by

$$Z_0 = \sqrt{\frac{(r + i\omega m)\tau}{i\omega}}, \qquad \Gamma = \sqrt{(r + i\omega m)\frac{i\omega}{\tau}}.$$

In the non-dissipative case we have simply

$$Z_0 = \sqrt{m\tau}, \qquad \Gamma = i\omega\sqrt{\frac{m}{\tau}}.$$

Heat Waves

Transmission of heat waves is also a special case of the generalized transmission line theory. In the one-dimensional case we have

$$\frac{\partial T}{\partial x} = -\frac{v}{K}, \qquad \frac{\partial v}{\partial x} = -c\delta\frac{\partial T}{\partial t},$$

where: T is the temperature, v the rate of heat flow, K the thermal conductivity, δ the density and c the specific heat. For simply periodic waves, we obtain

$$\frac{dT}{dx} = -\frac{1}{K}v, \qquad \frac{dv}{dx} = -i\omega c\delta\, T.$$

Thus the characteristic impedance and the propagation constant of heat waves are

$$Z_0 = \frac{1}{\sqrt{i\omega c\delta K}}, \qquad \Gamma = \sqrt{\frac{i\omega c\delta}{K}}.$$

The ratio "the temperature of the source/the rate of heat flow from the source" is the impedance "seen" by the heat source.

Electromagnetic Waves

The transmission equations of *uniform linearly polarized*[4] plane waves are:

$$\frac{dE}{dx} = -i\omega\mu H, \qquad \frac{dH}{dx} = -(g + i\omega\epsilon)E,$$

where: E is the electric intensity, H the magnetic intensity, and g, ϵ, μ are, respectively, the conductivity, the dielectric constant and permeability of the medium. These equations are of the same form as (2). Even the physical meanings of E and H are closely related to those of V and I; thus E is V per unit length and H is I per unit length.

The propagation constant and the characteristic impedance of an unbounded medium to linearly polarized plane waves are:

$$\sigma = \sqrt{i\omega\mu(g + i\omega\epsilon)}, \qquad \eta = \sqrt{\frac{i\omega\mu}{g + i\omega\epsilon}} = \frac{\sigma}{g + i\omega\epsilon} = \frac{i\omega\mu}{\sigma}.$$

These constants are so directly related to the fundamental electromagnetic constants of the medium that they themselves may be regarded as fundamental constants. On this account, we call σ and η, respectively, the *intrinsic propagation constant* and the *intrinsic impedance of the medium.* The intrinsic impedance will frequently occur as a multiplier in the expressions for the impedances of various types of waves.

The intrinsic impedance of a non-dissipative medium is simply $\eta = \sqrt{\mu/\epsilon}$; in air, this is equal to 120π or approximately 377 ohms.* Thus in the uniform linearly polarized plane wave traveling in free space, the relation between E and H is

$$E = 120\pi H \quad \text{or} \quad E \doteq 377H,$$

provided the positive directions of E and H are properly chosen.

An electromagnetic field of general character can be described by means of three electric components E_x, E_y, E_z, and three magnetic components H_x, H_y, H_z. We can form the following matrix whose components can be regarded as impedances:

[4] In this connection the word "uniform" is used to mean that equiphase planes are also equi-amplitude planes.

* See the letter from G. A. Campbell to Dean Harold Pender reproduced at the end of this paper.

$$\left\| \begin{array}{ccc} \frac{E_x}{H_x}, & \frac{E_x}{H_y}, & -\frac{E_x}{H_z} \\ -\frac{E_y}{H_x}, & \frac{E_y}{H_y}, & \frac{E_y}{H_z} \\ \frac{E_z}{H_x}, & -\frac{E_z}{H_y}, & \frac{E_z}{H_z} \end{array} \right\|$$

The algebraic signs preceding the ratios of components with different subscripts are assigned as follows. If a right-hand screw is rotated through 90° from the positive axis indicated by the subscript in the numerator toward the positive axis indicated by the subscript in the denominator, it will advance either in the positive or in the negative direction of the remaining axis. In the former case the ratio is given the positive sign and in the latter the negative sign. This convention happens to be particularly convenient in expressions for the Poynting vector.

Thus two impedances are associated with any pair of perpendicular directions, the x-axis and the y-axis, let us say; these impedances are:

$$Z_{xy} = \frac{E_x}{H_y}, \qquad Z_{yx} = -\frac{E_y}{H_x}.$$

If these two impedances are equal, then we define the *impedance in the direction of the positive z-axis* as follows:

$$Z_z = \frac{E_x}{H_y} = -\frac{E_y}{H_x}.$$

Similar definitions hold for the impedances in other directions.

While the impedances as now defined possess an attribute of direction, they are neither vectors nor tensors because they do not add in the proper fashion. However, in practical applications this lack of vectorial properties does not seem to be a drawback.

The above definitions can be extended to other systems of coordinates. Let r be the distance of a point P (r, θ, φ) from the origin of the spherical coordinate system, θ the polar angle or colatitude and φ the meridian angle or the longitude (Fig. 1). Then the "radial" impedance in the outward direction is defined as

$$Z_r = \frac{E_\theta}{H_\varphi} = -\frac{E_\varphi}{H_\theta}, \tag{6}$$

provided the two ratios of the field components are equal. The radial impedance looking *toward* the origin is defined as the negative of (6). Similarly the "meridian" impedance in the direction of increasing θ

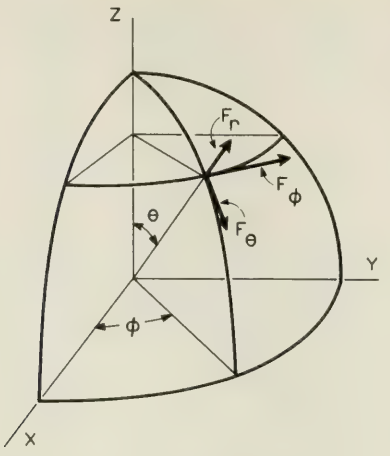


Fig. 1—Spherical coordinates. The positive directions of r-, θ-, and φ-components of a vector are, respectively, the directions of increasing r, θ, and φ.

and the impedance in the direction of increasing φ are:

$$Z_\theta = -\frac{E_r}{H_\varphi} = \frac{E_\varphi}{H_r}, \qquad Z_\varphi = \frac{E_r}{H_\theta} = -\frac{E_\theta}{H_r}.$$

In cylindrical coordinates we have (Fig. 2):

$$Z_\rho = -\frac{E_z}{H_\varphi} = \frac{E_\varphi}{H_z}, \qquad Z_\varphi = \frac{E_z}{H_\rho} = -\frac{E_\rho}{H_z}, \qquad Z_z = \frac{E_\rho}{H_\varphi} = -\frac{E_\varphi}{H_\rho}.$$

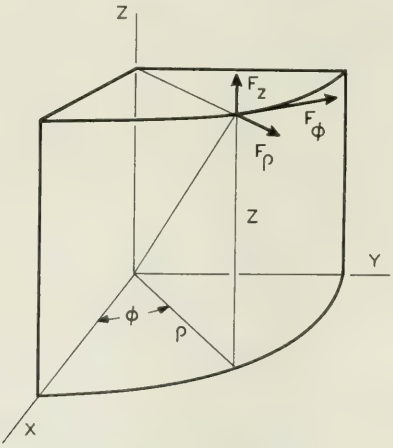


Fig. 2—Cylindrical coordinates. The positive directions of ρ-, φ-, and z-components of a vector are, respectively, the directions of increasing ρ, φ, and z.

Usually it is only one of the entire set of three-dimensional impedances that is of particular importance, the preferred direction being frequently the direction of the wave under consideration. When the ratios involved in the above definitions are unequal, it is expedient to resolve the field into component fields for which the ratios are equal. We shall now consider some special examples.

The field of the spherical electromagnetic wave emitted by a Hertzian doublet is known to be

$$\begin{aligned} E_\theta^+ &= \frac{i\omega\mu Ile^{-\sigma r}}{4\pi r}\left(1+\frac{1}{\sigma r}+\frac{1}{\sigma^2 r^2}\right)\sin\theta, \\ E_r^+ &= \frac{\eta Ile^{-\sigma r}}{2\pi r^2}\left(1+\frac{1}{\sigma r}\right)\cos\theta, \\ H_\varphi^+ &= \frac{\sigma Ile^{-\sigma r}}{4\pi r}\left(1+\frac{1}{\sigma r}\right)\sin\theta, \end{aligned} \tag{7}$$

where: Il is the moment of the doublet in ampere-meters, r is the distance from the doublet, θ the angle made by a typical direction in space with the axis of the doublet, and φ is the angle between two planes containing the doublet, one of which is kept fixed for reference. The radial impedance of this wave is

$$Z_r^+ = \frac{E_\theta}{H_\varphi} = \eta\,\frac{1+\frac{1}{\sigma r}+\frac{1}{\sigma^2 r^2}}{1+\frac{1}{\sigma r}}.$$

In a non-dissipative medium this becomes

$$Z_r^+ = \eta\,\frac{1+\frac{1}{i\beta r}-\frac{1}{\beta^2 r^2}}{1+\frac{1}{i\beta r}}.$$

At a distance large compared with the wave-length, the radial impedance to the spherical wave emitted by a doublet is substantially equal to the intrinsic impedance of the medium. Very close to the doublet (compared with the wave-length) the radial impedance is substantially a capacitive reactance; in fact, we have approximately $Z_0^+ = 1/i\omega\epsilon r$.

Reversing the sign of σ in (7), we obtain a spherical wave traveling toward the origin. At first sight, this *inward bound* wave appears to be the natural mate to the outward bound wave. Two such waves move in opposite directions in the same sense in which two plane

waves spread out from a plane source in an infinite homogeneous space. However, the analogy is not complete. The inward bound spherical wave cannot exist without an appropriate receiver of energy at the origin. In absence of such a receiver, the energy condenses at the origin and spreads outward again. The result of interference between two such progressive waves will be called the "internal" spherical wave.[5] It is natural to regard a thin spherical source in an infinite *homogeneous* medium as an analogue of a thin plane source and to consider the waves *on the two sides* of such a spherical source as the mates. In accordance with this idea the (+) and the (−) signs are used to distinguish between the waves produced by a source on its two sides rather than to indicate "progressive" waves moving in opposite directions. This attitude is not only a possible and a natural attitude but almost a necessary one in view of the fact that no generally applicable criterion is known by which "progressive" waves could be identified in any particular case. As often happens, in simple situations there is no need for arguing as to which attitude is the more proper one; thus the waves on the two sides of a plane source in an infinite homogeneous medium are two progressive waves moving in opposite directions.

The field of the internal spherical wave is

$$E_\theta^- = \frac{i\omega\mu A}{2\pi r}\left(\sinh \sigma r - \frac{\cosh \sigma r}{\sigma r} + \frac{\sinh \sigma r}{\sigma^2 r^2}\right)\sin\theta,$$

$$E_r^- = \frac{\eta A}{\pi r^2}\left(\frac{\sinh \sigma r}{\sigma r} - \cosh \sigma r\right)\cos\theta,$$

$$H_\varphi^- = \frac{\sigma A}{2\pi r}\left(\frac{\sinh \sigma r}{\sigma r} - \cosh \sigma r\right)\sin\theta.$$

The corresponding impedance is then

$$Z_r^- = -\frac{E_\theta^-}{H_\varphi^-} = \eta\,\frac{\sinh \sigma r - \dfrac{\cosh \sigma r}{\sigma r} + \dfrac{\sinh \sigma r}{\sigma^2 r^2}}{\cosh \sigma r - \dfrac{\sinh \sigma r}{\sigma r}}.$$

Close to the origin we have approximately

$$Z_r^- = \frac{2}{(g + i\omega\epsilon)r}.$$

[5] If the medium is non-dissipative, this wave is a *standing* wave; but, in general, it is simply a combination of two progressive waves in such proportions that the field is finite at the origin.

If the source of electromagnetic waves is a small coil rather than a small doublet, the field is

$$E_\varphi^+ = -\frac{\eta\sigma^2 SIe^{-\sigma r}}{4\pi r}\left(1+\frac{1}{\sigma r}\right)\sin\theta,$$
$$H_\theta^+ = \frac{\sigma^2 SIe^{-\sigma r}}{4\pi r}\left(1+\frac{1}{\sigma r}+\frac{1}{\sigma^2 r^2}\right)\sin\theta, \tag{8}$$
$$H_r^+ = \frac{\sigma SIe^{-\sigma r}}{2\pi r^2}\left(1+\frac{1}{\sigma r}\right)\cos\theta.$$

In this equation I is the current in the loop and S is the area. The corresponding radial impedance is then:

$$Z_r^+ = -\frac{E_\varphi^+}{H_\theta^+} = \eta\,\frac{1+\dfrac{1}{\sigma r}}{1+\dfrac{1}{\sigma r}+\dfrac{1}{\sigma^2 r^2}}.$$

This impedance approaches η as r increases indefinitely. Close to the loop we have approximately

$$Z_r^+ = i\omega\mu r. \tag{9}$$

The field of the internal wave having the same type of amplitude distribution over equiphase surfaces as the diverging wave (8) is

$$E_\varphi^- = \frac{\eta\sigma^2 A}{2\pi r}\left(\cosh\sigma r - \frac{\sinh\sigma r}{\sigma r}\right)\sin\theta,$$
$$H_\theta^- = \frac{\sigma^2 A}{2\pi r}\left(\sinh\sigma r - \frac{\cosh\sigma r}{\sigma r}+\frac{\sinh\sigma r}{\sigma^2 r^2}\right)\sin\theta,$$
$$H_r^- = \frac{\sigma A}{\pi r^2}\left(\frac{\sinh\sigma r}{\sigma r} - \cosh\sigma r\right)\cos\theta.$$

The radial impedance to this wave is then

$$Z_r^- = \frac{E_\varphi^-}{H_\theta^-} = \eta\,\frac{\cosh\sigma r - \dfrac{\sinh\sigma r}{\sigma r}}{\sinh\sigma r - \dfrac{\cosh\sigma r}{\sigma r}+\dfrac{\sinh\sigma r}{\sigma^2 r^2}}.$$

Close to the origin we have approximately

$$Z_r^- = \tfrac{1}{2}i\omega\mu r. \tag{10}$$

A line doublet formed by two parallel electric current filaments produces a cylindrical wave. Close to the doublet (compared with

the wave-length) we have

$$H_\varphi^+ = \frac{Il}{2\pi\rho^2}\cos\varphi, \qquad H_\rho^+ = -\frac{Il}{2\pi\rho^2}\sin\varphi. \tag{11}$$

In this equation Il is the moment of the doublet per unit length, I being the current and l the distance between the filaments. These equations are well known in the elementary theory of electromagnetism. The electric field is obtainable from (11) with the aid of Faraday's law of electromagnetic induction. This field and the corresponding radial impedance are

$$E_z^+ = -\frac{i\omega\mu Il}{2\pi\rho}\cos\varphi, \qquad Z_\rho^+ = i\omega\mu\rho. \tag{12}$$

The exact field of the line doublet and the corresponding radial impedance are:

$$E_z^+ = -\frac{\eta\sigma^2 Il}{2\pi}K_1(\sigma\rho)\cos\varphi, \qquad H_\varphi^+ = -\frac{\sigma^2 Il}{2\pi}K_1'(\sigma\rho)\cos\varphi,$$

$$H_\rho^+ = -\frac{\sigma Il}{2\pi\rho}K_1(\sigma\rho)\sin\varphi, \qquad Z_\rho^+ = -\eta\frac{K_1(\sigma\rho)}{K_1'(\sigma\rho)}.$$

The internal cylindrical wave with the same relative amplitude distribution over equiphase surfaces as in the wave originated by the line doublet is [6]

$$E_z^- = i\omega\mu A\, I_1(\sigma\rho)\cos\varphi, \qquad H_\varphi^- = \sigma A\, I_1'(\sigma\rho)\cos\varphi,$$

$$H_\rho^- = \frac{A}{\rho}I_1(\sigma\rho)\sin\varphi, \qquad Z_\rho^- = \eta\frac{I_1(\sigma\rho)}{I_1'(\sigma\rho)}.$$

Close to the doublet we have approximately

$$E_z^- = i\omega\mu P\rho\cos\varphi, \qquad H_\varphi^- = P\cos\varphi,$$
$$H_\rho^- = P\sin\varphi, \qquad Z_\rho^- = i\omega\mu\rho.$$

Another familiar field is that produced by two parallel line charges in a perfect dielectric. Close to the doublet this field is

$$E_\varphi^+ = \frac{ql\sin\varphi}{2\pi\epsilon\rho^2}, \qquad E_\rho^+ = \frac{ql\cos\varphi}{2\pi\epsilon\rho^2},$$
$$H_z^+ = \frac{i\omega ql\sin\varphi}{2\pi\rho}, \qquad Z_\rho^+ = \frac{1}{i\omega\epsilon\rho}, \tag{13}$$

[6] The symbols $I_n(x)$ and $K_n(x)$ designate the modified Bessel functions as defined in G. N. Watson's "Bessel Functions."

ql being the moment of the doublet. The last equation is obtained from the first two with the aid of Ampére's law. The exact expressions for any medium are

$$E_\varphi^+ = -\frac{i\omega ql\sigma\eta}{2\pi} K_1'(\sigma\rho)\sin\varphi, \qquad E_\rho^+ = \frac{i\omega ql\eta}{2\pi\rho} K_1(\sigma\rho)\cos\varphi,$$

$$H_z^+ = \frac{i\omega ql\sigma}{2\pi} K_1(\sigma\rho)\sin\varphi, \qquad Z_\rho^+ = -\eta\frac{K_1'(\sigma\rho)}{K_1(\sigma\rho)}.$$

For an internal cylindrical wave, we have

$$E_\varphi^- = A\sigma I_1'(\sigma\rho)\sin\varphi, \qquad E_\rho^- = -\frac{A}{\rho} I_1(\sigma\rho)\cos\varphi,$$

$$H_z^- = -A(g + i\omega\epsilon) I_1(\sigma\rho)\sin\varphi, \qquad Z_\rho^- = \eta\frac{I_1'(\sigma\rho)}{I_1(\sigma\rho)}.$$

Close to the doublet this becomes substantially

$$E_\varphi^- = P\sin\varphi, \qquad E_\rho^- = -P\cos\varphi,$$

$$H_z^- = -P(g + i\omega\epsilon)\rho\sin\varphi, \qquad Z_\rho^- = \frac{1}{(g + i\omega\epsilon)\rho}.$$

In concluding this set of examples we shall emphasize the fact that the impedance to a wave depends upon the particular manner in which the applied electromotive force is distributed in space, in very much the same way as it depends upon the manner of distribution of this force in time, that is, upon the frequency of the wave. Just as the impedance has a meaning only if the applied electromotive force varies harmonically with a certain well defined frequency,[7] there are definite types of applied force distribution in space for which the impedance has a meaning and other types for which it has not. Arbitrary spatial distributions of force may be decomposed into "space harmonics" in a manner analogous to Fourier's frequency analysis of arbitrary time distributions of force. This is just another way of interpreting the well-known method of solving Maxwell's equations with the aid of characteristic wave functions.

Here is a simple example of the dependence of the impedance to a wave upon the manner of applied force distribution. Consider the wave generated by an infinite electric current filament of radius a

[7] Strictly speaking, the impedance concept is applicable to any impressed force which varies exponentially with time, the exponent being in general a complex number. The only exceptions are the exponents which are either zeros or infinities of the impedance function. Undamped impressed forces constitute merely an important subclass of exponential forces.

when the electromotive force driving the current is distributed uniformly along the filament. In this case we have

$$E_z = -\frac{\eta I K_0(\sigma\rho)}{2\pi a K_1(\sigma a)}, \qquad H_\varphi = \frac{I K_1(\sigma\rho)}{2\pi a K_1(\sigma a)}, \qquad Z_\rho = \eta \frac{K_0(\sigma\rho)}{K_1(\sigma\rho)},$$

where I is the current in the filament. On the other hand if the electromotive force is applied to the filament with a uniform progressive phase delay so that it varies along the filament as e^{-ikz} for instance, then the field and the impedance are

$$E_z = -\frac{\eta I K_0(\Gamma\rho)}{2\pi a K_1(\Gamma a)} e^{-ikz}, \qquad H_\varphi = \frac{I K_1(\Gamma\rho)}{2\pi a K_1(\Gamma a)} e^{-ikz},$$

$$Z_\rho = \eta \frac{K_0(\Gamma\rho)}{K_1(\Gamma\rho)}, \qquad \Gamma = \sqrt{\sigma^2 + k^2}.$$

PART II

REFLECTION, REFRACTION, SHIELDING AND POWER ABSORPTION—GENERAL FORMULÆ

Uniform Transmission Lines

While the following discussion refers specifically to an electric transmission line, the results apply to all generalized transmission lines of which the former may be considered typical. These results depend upon certain boundary conditions and are not influenced by the names of the variables.

Consider a semi-infinite transmission line terminated by a prescribed impedance Z_t. Suppose that an "impressed" wave is coming from infinity. If V_i and I_i are the voltage and the current, their ratio must equal the characteristic impedance Z_0 of the wave. On the other hand, the ratio of the voltage across the impedance Z_t to the current through it is Z_t by definition. Thus, unless Z_t is equal to Z_0, a "reflected" wave must originate at the terminal and travel backwards. Let V_r and I_r be the voltage and the current of the reflected wave at the terminal. The total values of the voltage and the current will be designated by V_t and I_t. Then at the terminal

$$I_i + I_r = I_t, \qquad V_i + V_r = V_t. \tag{14}$$

By (9) and by the definition of Z_t, we have

$$V_i = Z_0 I_i, \qquad V_r = -Z_0 I_r, \qquad V_t = Z_t I_t. \tag{15}$$

Designating the ratio of the characteristic impedance of the line to

the terminal impedance by k, we use (15) to rewrite (14) in the following form:

$$I_i + I_r = I_t, \qquad I_i - I_r = \frac{I_t}{k}, \qquad k = \frac{Z_0}{Z_t}.$$

Solving, we obtain the reflection and transmission coefficients:

$$\begin{aligned} R_I &= \frac{I_r}{I_i} = \frac{k-1}{k+1}, & R_V &= \frac{V_r}{V_i} = \frac{1-k}{1+k}, \\ T_I &= \frac{I_t}{I_i} = \frac{2k}{1+k}, & T_V &= \frac{V_t}{V_i} = \frac{2}{1+k}. \end{aligned} \tag{16}$$

Thus *when $k = 1$, that is when the terminal impedance equals the characteristic impedance, there is no reflection. When the ratio of the impedances is zero or infinity the reflection is complete: in the first case the current vanishes and the voltage is doubled, and in the second the current is doubled and the voltage vanishes. The amount of reflection is completely determined by the ratio of the impedances.*

The terminal impedance may be another semi-infinite transmission line and its characteristic impedance will play the part of Z_t. It is important to note that *neither the propagation constants nor the velocities of the wave in the lines have anything to do with reflection.* No reflection will take place if the lines have equal impedances and there will be reflection in the case of unequal impedances even if the velocities are the same.

The variables V and I can stand for any two physical quantities satisfying equations (2). It will be observed that if we disregard the physical significance of the variables V and I, the characteristic impedance can be defined either as the ratio V/I or as I/V. We are perfectly free to make our choice. It is evident from (16) that if we interchange V and I and replace k by its reciprocal, the expressions for the reflection and transmission coefficients remain unaltered.

Non-Uniform Transmission Lines

The foregoing analysis has to do only with *uniform* lines. In the case of non-uniform lines the impedances looking in the opposite directions may be different. These two impedances will be defined by the following equations:

$$Z_0^+ = \frac{V^+}{I^+}, \qquad Z_0^- = -\frac{V^-}{I^-},$$

where V^+, I^+ and V^-, I^- refer to the two waves.

At the terminal we have as before:

$$I_i + I_r = I_t, \qquad Z_0^+ I_i - Z_0^- I_r = Z_t I_t.$$

Hence the more general expressions for the reflection and transmission coefficients are:

$$R_I = \frac{Z_0^+ - Z_t}{Z_0^- + Z_t}, \qquad R_V = -\frac{Z_0^-}{Z_0^+} R_I,$$

$$T_I = \frac{Z_0^- + Z_0^+}{Z_0^- + Z_t}, \qquad T_V = \frac{Z_t}{Z_0^+} T_I.$$

These reflection and transmission coefficients can be expressed in terms of the ratios of the line impedances to the terminal impedance.

Shielding

When a source of electromagnetic waves is enclosed in a metallic box, the field outside the box is substantially weaker than it would have been in the absence of the box. The box is said to act as a "shield." Under some conditions, transmission of electromagnetic waves in free space and in the metallic shield is governed by equations of the form (1). In those cases the shielding effect can evidently be regarded as due to a reflection loss at the boundaries of the shield and to an attenuation loss in the shield itself. A schematic representation of a single layer shield is shown in Fig. 3. The source of disturbance

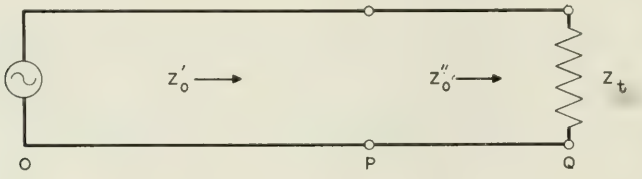


Fig. 3—Transmission line representation of a shield. The generator represents the source of the electromagnetic disturbance, the section OP the space surrounding the source, the section PQ the shield, and the impedance Z_t the space outside the shield.

is shown as a generator, the space around this source is represented by a piece of a transmission line OP, the shield by a piece PQ and the space outside the shield by the impedance Z_t.

The simplest case to consider is that of an *electrically thick* shield, in which the attenuation between P and Q is so great that waves reflected at Q do not affect appreciably the situation at P. In such a case the impedance at P looking toward Q equals the characteristic impedance Z_0'' and the same is true of the impedance at Q looking

toward P. The effect of the inserted piece is comprised of two independent reflections at P and Q and of attenuation with concomitant phase change between P and Q. Thus the transmission coefficients across PQ, that is, the ratios of the quantities at Q to the impressed quantities at P, are

$$T_I = T_{I,P}\, T_{I,Q}\, e^{-\Gamma'' l}, \qquad T_V = T_{V,P}\, T_{V,Q}\, e^{-\Gamma'' l},$$

where $T_{I,P}$ is the transmission coefficient for I at P and the remaining T's have similar meanings.

If PQ is a piece of a uniform transmission line inserted into a uniform semi-infinite line, $Z_t = Z_0'$. In this case, we have

$$T_I = T_V = \frac{4k}{(k+1)^2}\, e^{-\Gamma'' l},$$

where k is the ratio of the characteristic impedances. The factor $4k/(k+1)^2$ represents the reflection loss and $e^{-\alpha'' l}$ the attenuation loss.

Let us now assume that PQ is electrically short and that all the transmission lines in question are uniform. By the transmission line theory, the ratios of the total currents and voltages at P and Q are:

$$\frac{I_Q}{I_P} = \frac{Z_0''}{Z_0'' \cosh \Gamma'' l + Z_t \sinh \Gamma'' l},$$

$$\frac{V_Q}{V_P} = \frac{Z_t}{Z_t \cosh \Gamma'' l + Z_0'' \sinh \Gamma'' l}.$$

On the other hand, we have

$$\frac{I_P}{I_i} = \frac{2Z_0'}{Z_0' + Z_P}, \qquad \frac{V_P}{V_i} = \frac{2Z_P}{Z_0' + Z_P},$$

where Z_P is the impedance at P looking toward Q

$$Z_P = Z_0'' \frac{Z_t \cosh \Gamma'' l + Z_0'' \sinh \Gamma'' l}{Z_0'' \cosh \Gamma'' l + Z_t \sinh \Gamma'' l}.$$

The transmission coefficients across PQ can be represented as

$$T_I = \frac{I_Q}{I_i} = \frac{I_Q}{I_P} \cdot \frac{I_P}{I_i}, \qquad T_V = \frac{V_Q}{V_i} = \frac{V_Q}{V_P} \cdot \frac{V_P}{V_i}.$$

Making appropriate substitutions into this equation, we obtain

$$T_I = p(1 - qe^{-2\Gamma'' l})^{-1} e^{-\Gamma'' l}, \qquad T_V = \frac{Z_t}{Z_0'}\, T_I, \tag{17}$$

where

$$p = \frac{4Z_0'Z_0''}{(Z_0'' + Z_0')(Z_0'' + Z_t)},$$

$$q = \frac{(Z_0'' - Z_0')(Z_0'' - Z_t)}{(Z_0'' + Z_0')(Z_0'' + Z_t)}.$$

In the special case when $Z_t = Z_0'$, we have

$$p = \frac{4k}{(k+1)^2}, \qquad q = \left(\frac{k-1}{k+1}\right)^2, \qquad \text{and} \qquad T_V = T_I.$$

If PQ is electrically long, (17) becomes simply

$$T_I = pe^{-\Gamma'' l}, \qquad T_V = \frac{Z_t}{Z_0'} T_I.$$

An interesting physical interpretation of (17) will follow if we expand the factor in parentheses into a series

$$T_I = pe^{-\Gamma'' l} + pqe^{-3\Gamma'' l} + pq^2e^{-5\Gamma'' l} + \cdots.$$

The first term represents what remains of the original wave on the first passage through PQ. A part of the original wave is reflected back at Q and then partially re-reflected from P; the second term represents that fraction of the re-reflected wave which is transmitted beyond Q. The following terms represent succeeding reflections. In making this analysis, we must remember that $p = p_1p_2$ where p_1 and p_2 are respectively the transmission coefficients across the first and the second boundaries on the supposition that the inserted piece is infinitely long. Similarly, $q = q_1q_2$, the product of the two reflection coefficients.

Let us now consider a non-uniform transmission line. The propagation of a disturbance is no longer exponential and we introduce the ratios $\kappa^+ = V^+(x_2)/V^+(x_1)$ and $\kappa^- = V^-(x_1)/V^-(x_2)$ for the voltage ratios in the waves moving in opposite directions. In what follows x_1 and x_2 are the coordinates of the beginning and the end of the inserted piece. The transmission coefficient T across the insertion, that is, the ratio of the total quantity at $x = x_2$ to the impressed quantity at $x = x_1$, is then

$$T = p_1\kappa^+p_2 + (p_1\kappa^+)(q_2\kappa^-q_1\kappa^+)p_2 + (p_1\kappa^+)(q_2\kappa^-q_1\kappa^+)(q_2\kappa^-q_1\kappa^+)p_2 + \cdots.$$

This can be rewritten as follows:

$$T = p[1 + q\kappa + (q\kappa)^2 + (q\kappa)^3 + \cdots]\kappa^+ = \frac{p}{1 - q\kappa}\kappa^+,$$

where

$$p = p_1 p_2, \qquad q = q_1 q_2, \qquad \kappa = \kappa^+ \kappa^-.$$

The same formula applies of course to I provided we interpret the p's, q's and κ's as referring to the variable I rather than the variable V. If the inserted piece is electrically long, we have approximately

$$T = p\kappa^+.$$

In many practical applications the inserted piece is a uniform transmission line, so that

$$\kappa^+ = \kappa^- = e^{-\Gamma l}, \qquad \kappa = e^{-2\Gamma l},$$

where Γ is the propagation constant and l is the length of the piece. In this case

$$T = \frac{p}{1 - qe^{-2\Gamma l}} e^{-\Gamma l}.$$

Power Absorption and Radiation

The power transferred from left to right across P is the real part of the following function Ψ:

$$\Psi_P = \tfrac{1}{2} V_P I_P^* = \tfrac{1}{2} Z_P I_P I_P^*, \tag{18}$$

where the asterisk denotes the complex number conjugate to the one represented by the letter itself. The power absorbed by the impedance Z_t is

$$\Psi_T = \tfrac{1}{2} Z_t I_Q I_Q^*. \tag{19}$$

The difference between (18) and (19) represents the power absorbed by the section PQ.

The power absorbed by a shield is calculated in a similar manner. The energy flow per unit area of the shield is given by an expression closely analogous to (18); the tangential component of H appears in the place of I and Z_P is to be interpreted as the impedance in the direction normal to the shield. The formula is derivable from the Poynting expression for energy flow. Thus the power flow per unit area is $\frac{1}{2} Z_n H_t H_t^*$ where H_t is the tangential component of H and Z_n is the impedance in the direction normal to the shield.

PART III

REFLECTION, REFRACTION, SHIELDING AND POWER ABSORPTION

Special Applications

The general formulæ derived in the preceding part are directly applicable to a variety of special cases such as reflection of plane waves at a plane boundary, shielding action of cylindrical shields upon electric waves produced by an infinite parallel pair of electric current filaments, shielding action of spherical shields upon electric waves produced by a coil or a condenser, etc. Of course, most of these results have already been obtained and published, each special problem having been treated on its own merits rather than as a particular case of a general formula. For this reason, we shall confine ourselves largely to a discussion of those aspects of reflection which are particularly illuminated by the general point of view.

Cylindrical Waves

Consider two parallel wires carrying equal and opposite alternating currents. At a distance from the wires two or three times as large as their interaxial separation, the wave is substantially that of a line doublet and the radial impedance in free space is approximately [8] $i\omega\mu\rho$ so long as ρ is much less than the wave-length. This restriction on ρ is permissible in the present communication art. In metallic media this expression for the radial impedance is good only at very low frequencies. At high frequencies the radial impedance in metallic media is substantially [9] $\sqrt{i\omega\mu/g}$.

If the pair of wires is surrounded by a metal cylinder, the latter will act as a shield by virtue of reflections taking place at the boundary and attenuation through the shield.

The attenuation constant is substantially $\sqrt{\pi\mu gf}$ nepers per meter.[10] Thus the attenuation in logarithmic units through the shield is proportional to the first power of its thickness and to the square roots of the conductivity, the permeability and the frequency.

The reflection loss depends upon the impedance ratio. In the neighborhood of $f = 0$, the impedance ratio is seen to be equal to the ratio of the permeabilities. Consequently, at very low frequencies non-magnetic shields are relatively inefficient since there is no reflec-

[8] Equation (12).

[9] The approximate error is $1/2\sigma\rho$; at 10 kc. the error is about 2.5 per cent at a distance 1 cm. from the line source.

[10] This is true even at low frequencies if the shield is thin compared to its diameter. Otherwise, the cylindrical divergence of the wave must be taken into account.

tion loss. Inasmuch as the radial impedance in air is proportional to the first power of the frequency and in metal it is proportional only to the square root of the frequency, a point is reached beyond which the radial impedance in air always exceeds the radial impedance in metals. Thus the air-to-magnetic metal impedance ratio is less than unity near $f = 0$ and greater than unity for sufficiently high frequencies. Consequently, the absolute value of this impedance ratio is equal to unity at some intermediate frequency at which the reflection

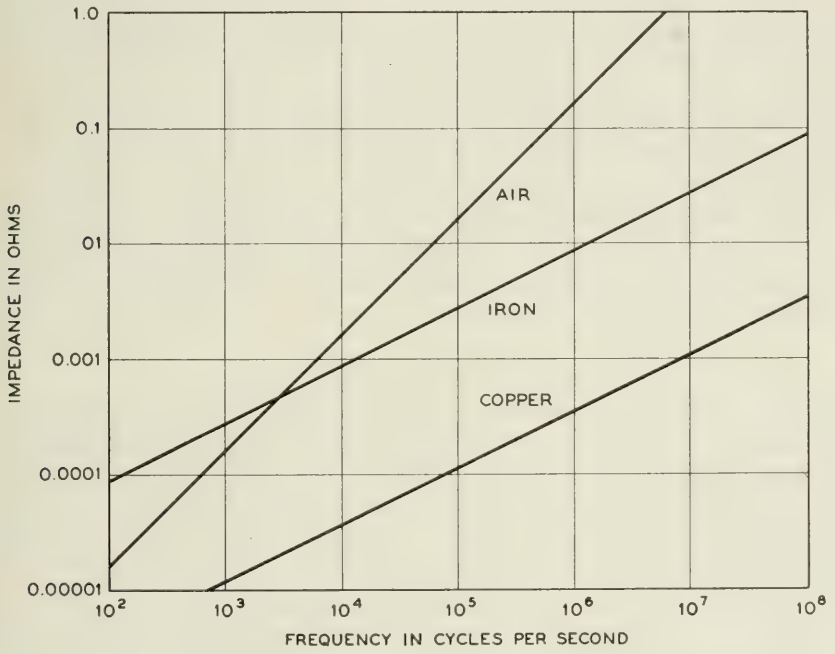


Fig. 4—The radial impedances in air, copper and iron at a distance of 2 centimeters from the axis for cylindrical waves generated by line doublets comprised of infinitely long electric current filaments. The conductivity of copper = 5.8005×10^7 mhos per meter, the conductivity of iron = 10^7 mhos per meter, the permeability of air and copper = 1.257×10^{-6} henries per meter, the permeability of iron = 1.257×10^{-4} henries per meter.

loss will be quite small.[11] Some typical curves of radial impedances are shown in Fig. 4. The radial impedances in non-magnetic metals are always less than the impedance in air.

At high frequencies the reflection loss between metals is substantially independent of the frequency. At copper-iron boundaries this loss is always high and at copper-air boundaries it increases steadily with the frequency and becomes quite substantial at frequencies as

[11] A small reflection loss exists because the impedances have different phases.

high as 100,000 cycles. On the other hand, in a certain frequency range the reflection loss at iron-air boundaries may be very low. Since the attenuation loss of a complete shield made of coaxial layers of copper and iron is independent of the sequence of the layers, considerable gain in shielding may be secured by placing an iron layer between two copper layers rather than a copper layer between two iron layers so as to take advantage of the added reflection loss, assuming of course that the amounts of copper and iron are the same in both cases.

Since the high-frequency impedance ratio is proportional to the diameter of the shield, the size has a substantial influence upon the effectiveness of the shield. Each time the diameter of a *non-magnetic* shield is doubled, the shielding is increased by 6 decibels. In the case of *magnetic* shields, this is true only at frequencies considerably higher than the critical frequency at which the reflection loss is minimum. Considerably below this frequency, the effectiveness of a magnetic shield is *decreased* by 6 decibels with each doubling of the diameter of the shield. For the transition region we can say that with increasing size the effectiveness of the magnetic shield decreases below the critical frequency and increases above it.

"Electrostatic Shielding"

If the cylindrical wave is originated by two parallel oppositely charged wires, alternating with a given frequency f, the radial impedance in free space is $1/i\omega\epsilon\rho$ provided ρ is small compared with the wave-length.[12] As in the preceding case, in metallic media the radial impedance is $\sqrt{i\omega\mu/g}$ provided the frequency is not too low; for very low frequencies the radial impedance becomes $1/g\rho$.

It is clear at once that for these waves the reflection loss is tremendous. Thus in air $\epsilon = (1/36\pi)10^{-9}$ farads per meter; if $f = 10^6$ and $\rho = 0.01$ m., then the radial impedance is 1,800,000 ohms. The corresponding impedance in copper is only 0.000369 ohms. At lower frequencies the disparity between the radial impedances becomes even greater. The impedance ratio tends to infinity as the frequency approaches zero.

In the elementary theory a metal shield is regarded as a perfect shield against this "electrostatic field." An "electrostatic" field alternating 1,000,000 cycles per second is probably a misnomer. And the shielding is excellent but not perfect. Nevertheless the distinction between two possible types of waves is a valid one, at least in the frequency range usually employed in the communication art. In one wave the electric field is normal to the direction of propagation and

[12] Equation (13).

in the other the magnetic field is so disposed. The former wave may be called *transverse electric* and the latter *transverse magnetic*. The product of the corresponding radial impedances of these waves is equal to the square of the intrinsic impedance. Hence if one wave is a low impedance wave (as compared to the intrinsic impedance), the other is a high impedance wave. Under the usual engineering conditions these waves are unmistakably different in air, although this distinction disappears in metallic media. It must be pointed out, however, that for micro-waves the dimensions of the shield may be comparable to the wave-length, in which case the radial impedances may be of the same order of magnitude.

In the above discussion we have supposed that the line source was on the axis of the shield. If it is not, it is possible to represent the actual source by means of an equivalent system of sources along the axis and calculate the shielding effect. The latter is different for cylindrical waves of different orders. This will result in somewhat different shielding for different positions outside the shield. Ordinarily, however, the difference is not large enough to be considered in practical problems.

Spherical Waves

A small coil carrying an alternating current will give rise to a transverse electric spherical wave and a small condenser to a transverse magnetic wave. Consider a shield concentric with the coil or the condenser. In the shield the radial impedance is $\sqrt{i\omega\mu/g}$, again excepting very low frequencies. In air the radial impedance of the outward bound electric wave is [13] $i\omega\mu r$ and that of the internal wave $\frac{1}{2}i\omega\mu r$. The corresponding impedances of transverse magnetic waves are $1/i\omega\epsilon r$ and $2/i\omega\epsilon r$. The conditions for reflection and shielding are substantially the same as in the case of cylindrical waves. Some quantitative difference results from the inequality of the radial impedances in opposite directions.

Plane Waves

The next example of uniform linearly polarized plane waves is particularly well known.[14] When the boundary between two media coincides with an equiphase surface of the impinging wave, the formulæ

[13] Equations (9), (10).

[14] The general formulæ for the reflection and transmission coefficients have been obtained by T. C. Fry on the basis of the Maxwell theory in his paper "Plane Waves of Light II," published in the *Journal of the Optical Society of America* and *Review of Scientific Instruments*, Vol. 16, pp. 1–25 (1928). The earliest formulæ are probably due to A. Cauchy, who obtained them from the "elastic solid" theory of light waves.

of Part II are directly applicable and the reflection coefficient depends upon the ratio of the intrinsic impedances of the media.

A more interesting situation arises when the incidence is oblique. Let the xy-plane be the boundary between two homogeneous media and let the *electric vector be parallel* to this boundary. We may assume it to be parallel to the x-axis. In this case the electric field strength is given by

$$E_x = E_0 e^{-\sigma s + i\omega t}, \qquad E_y = E_z = 0, \tag{20}$$

where E_0 is the amplitude and s is the distance from the equiphase surface passing through the origin. If the angle of incidence is ϑ (Fig. 5), this distance may be expressed as: $s = y \sin \vartheta + z \cos \vartheta$.

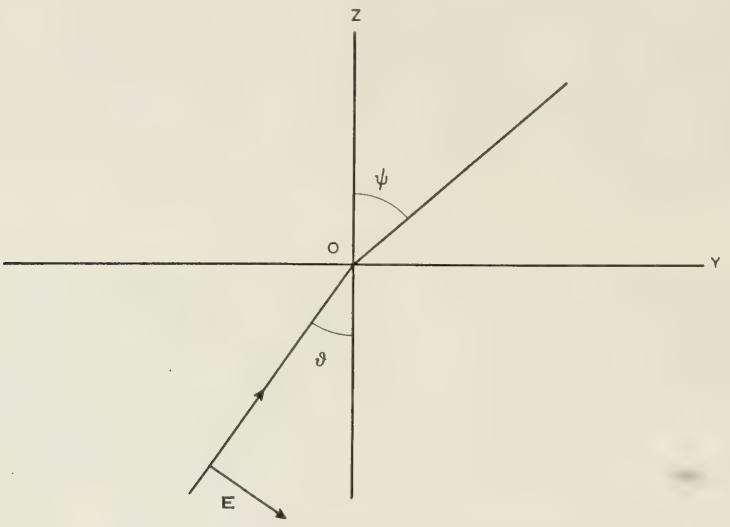


Fig. 5—Reflection of plane waves. The x-axis is toward the reader, the xy-plane is the boundary between the media, the E-vector is toward the reader, the angle of incidence = ϑ, and the angle of reflection = ψ.

The magnetic vector is perpendicular to E and to the ray and its value is

$$H = H_0 e^{-\sigma s + i\omega t}, \qquad E_0 = \eta H_0. \tag{21}$$

The cartesian components are then

$$H_y = H_0 \cos \vartheta \, e^{-\sigma s + i\omega t}, \qquad H_z = -H_0 \sin \vartheta \, e^{-\sigma s + i\omega t}, \qquad H_x = 0.$$

Equations (20) and (21) represent the motion of equiphase planes in the direction specified by the angle ϑ. It is equally possible to regard them as representing the motion of *phase-amplitude patterns*

in the direction *normal* to the xy-plane. We need only to rewrite these equations as follows:

$$\begin{aligned} E_x &= (E_0 e^{-\sigma y \sin \vartheta}) e^{-\sigma z \cos \vartheta + i\omega t}, \\ H_y &= (H_0 \cos \vartheta \, e^{-\sigma y \sin \vartheta}) e^{-\sigma z \cos \vartheta + i\omega t}. \end{aligned} \tag{22}$$

The relative distribution of the amplitude and the phase of the wave are governed by the factor $e^{-\sigma y \sin \vartheta}$ and this phase-amplitude pattern is propagated in the direction of the z-axis, the propagation constant being $\sigma \cos \vartheta$.

The advantages of this point of view are clear. In attempting to find the reaction of the second medium upon the incident wave, it is necessary to satisfy certain boundary conditions at *every point* of the interface. This can be insured by requiring the reflected and the refracted waves to have the same phase-amplitude patterns at the interface and by adjusting their relative amplitude and phases to secure the fulfilment of the boundary conditions at some one point. In other words, the problem is reduced to that for which the general solution was given in Part II.

The impedance to the incident wave in the z-direction is found from (22):

$$Z_z = \frac{E_x}{H_y} = \frac{E_0}{H_0 \cos \vartheta} = \eta \sec \vartheta.$$

This impedance is seen to be a function of the intrinsic impedance of the medium and of the angle of incidence.

For the refracted wave in the second medium the transmission equations are similar to (22):

$$\begin{aligned} E_x' &= (E_0' e^{-\sigma' y \sin \psi}) e^{-\sigma' z \cos \psi + i\omega t}, \\ H_y' &= (H_0' \cos \psi \, e^{-\sigma' y \sin \psi}) e^{-\sigma' z \cos \psi + i\omega t}, \qquad E_0' = \eta' H_0. \end{aligned} \tag{23}$$

The "angle of refraction" ψ is, in general, different from ϑ. In our equations we may regard ψ merely as a parameter. Its value is obtained from the condition that at the xy-plane the phase-amplitude pattern of the incident and the refracted waves must be the same, and consequently

$$\sigma \sin \vartheta = \sigma' \sin \psi. \tag{24}$$

In dielectrics this relation is known as Snell's law of refraction.

By (23), the impedance to the refracted wave in the z-direction is

$$Z_z' = \frac{E_x'}{H_y'} = \eta' \sec \psi.$$

The reflection and the transmission coefficients are then obtained from (22) in terms of the impedance ratio

$$k = \frac{\eta \sec \vartheta}{\eta' \sec \psi} = \frac{\eta \cos \psi}{\eta' \cos \vartheta}. \tag{25}$$

Thus, we have

$$R_H = \frac{k - 1}{k + 1}, \qquad R_E = \frac{1 - k}{1 + k},$$

$$T_H = \frac{2k}{k + 1}, \qquad T_E = \frac{2}{1 + k}.$$

These coefficients refer to the *tangential* components of the field.

In a similar way we can deal with the case in which the magnetic vector of the incident wave is parallel to the boundary. The parts played by E and H are interchanged and the impedance ratio becomes

$$k = \frac{\eta \cos \vartheta}{\eta' \cos \psi}. \tag{26}$$

The cosine factors have changed their places.

The general case, in which neither E nor H is parallel to the boundary, cannot be treated in the above manner. In this case the components of E and H which are parallel to the boundary are not perpendicular to each other, the impedances Z_{xy} and Z_{yx} are not equal to each other and the unique impedance $Z_z = Z_{xy} = Z_{yx}$, upon which the results of Part II are based, does not exist. In accordance with a suggestion made in Part I, the incident wave must be resolved into components possessing unique impedances in the direction normal to the boundary. It is well known that such a decomposition is possible for ordinary plane waves; the latter can always be decomposed into two components, in one of which E is parallel to the boundary and in the other H is so disposed.

It is not surprising that reflection of arbitrarily oriented waves cannot be treated directly. The impedance ratios (25) and (26) for two basic orientations are in general different and the polarization of the reflected wave will be changed. An exceptional case arises *when the intrinsic propagation constants of the media are equal.* In this case $\psi = \vartheta$, as seen from (24), and *the impedance ratio is independent of the angle of incidence and of the particular orientation of the wave.* Consequently, *the reflection and the transmission coefficients depend solely upon the ratio of the intrinsic impedances of the media.*

Frequently the permeabilities of the media are assumed to be the same, in which case the ratio of the intrinsic impedances is equal to

the inverse ratio of the "indices of refraction" of the media. Much could be said, however, in favor of not making such an assumption when formulating the general results since in many applications the permeabilities may be unequal.

Images

A few additional interesting results can be obtained for the special case of two semi-infinite homogeneous media having equal propagation constants. If the media are separated by a plane boundary, problems of reflection and refraction can be solved by the *method of images*. This method is frequently used in electrostatics and one or

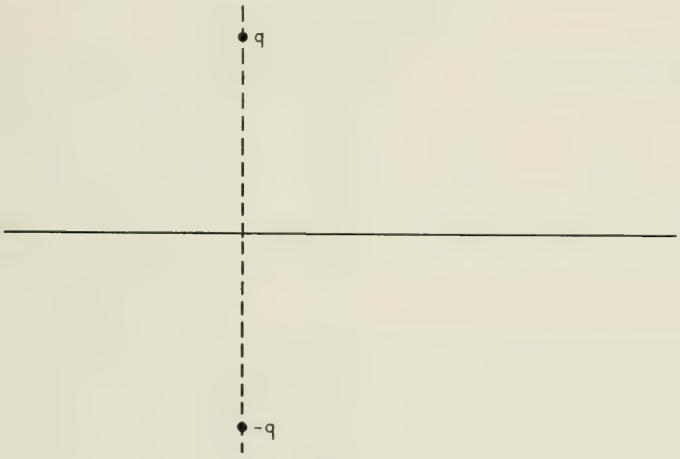


Fig. 6

two simple examples from that science will serve as an introduction to the later generalizations.

The field of a point charge q above a conducting plane can be found by assuming another point charge $(-q)$. This "image" charge (Fig. 6) is the same distance below the plane as the actual charge is above the plane, both charges lying on the same perpendicular. The field due to the original charge and to the image charge satisfies the boundary conditions at the conducting plane since it makes the latter an equipotential. This combined field gives the correct resultant field on the *same* side of the plane as the original charge; on the opposite side the field is zero.

If the boundary is the interface between two perfect dielectrics (Fig. 7) with dielectric constants respectively equal to ϵ_1 and ϵ_2, the results are almost equally simple. Above the boundary we have

a reflected field in addition to the original field. This reflected field is produced by an image charge $q' = (\epsilon_1 - \epsilon_2)q/(\epsilon_1 + \epsilon_2)$ on the supposition that the dielectric constant is everywhere equal to ϵ_1. Below the plane the field is such as would be produced by a charge $q'' = 2\epsilon_1 q/(\epsilon_1 + \epsilon_2)$ if placed where the original charge is, also on the assumption that the dielectric constant is everywhere ϵ_1. The charge producing the correct field below the boundary would be $q''' = 2\epsilon_2 q/(\epsilon_1 + \epsilon_2)$ if we were to assume ϵ_2 as the dielectric constant of the whole space.

Inspecting equations (7) for an electric current element, which we assume to be perpendicular to the plane interface of two homogeneous

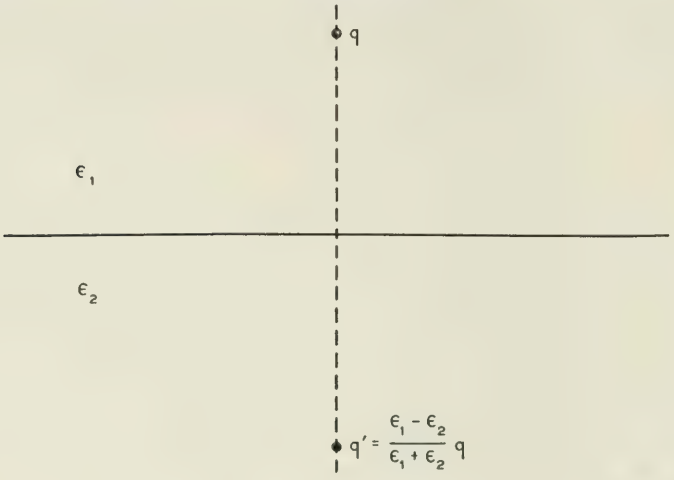


Fig. 7

media, we see that the method of images can readily be extended to *dynamic* fields *provided* the intrinsic propagation constants of the media are equal. In order to make this conclusion more evident, we replace $i\omega\mu$ in the first equation by the equivalent product $\eta\sigma$ and then calculate the component of E *tangential* to the interface

$$E_\rho^+ = E_\theta^+ \cos\theta + E_r^+ \sin\theta = \eta Il \frac{\sigma e^{-\sigma r}}{4\pi r}\left(1 + \frac{3}{\sigma r} + \frac{3}{\sigma^2 r^2}\right)\sin\theta\cos\theta.$$

It is easy to see that the continuity of the tangential field components will be preserved if we assume a reflected field on the same side of the boundary and a refracted field on the opposite side in accordance with the following specifications. *The reflected field is such as would be produced by an image current element of moment* $(\eta_1 - \eta_2)Il/(\eta_1 + \eta_2)$

and the refracted field is such as could be produced by a current element of moment $2\eta_1 Il/(\eta_1 + \eta_2)$, *occupying the same position as the source. In calculating these fields a uniform intrinsic impedance* η_1 *is assumed throughout the whole space.*

Since the current I in the element implies two point charges, $-I/i\omega$ and $I/i\omega$, at its terminals, we can interpret the above rule of images in terms of the charges. *The image of a point charge* q *for calculating the reflected field is* $(\eta_2 - \eta_1)q/(\eta_2 + \eta_1)$. *For calculating the refracted field a charge* $2\eta_1 q/(\eta_1 + \eta_2)$ *must be assumed in the same position as the original charge.* For perfect dielectrics the expressions of the image charges reduce to those given by electrostatics.

ACKNOWLEDGMENT

I wish to express my appreciation to Dr. T. C. Fry for his valuable criticism in the preparation of this paper.

AN HISTORICAL NOTE

The following memorandum written in 1932 by Dr. G. A. Campbell, formerly of the American Telephone and Telegraph Company, represents an interesting historical comment and it is reprinted with Dr. Campbell's permission.

A letter discussing the characteristic impedance of free space, written to me seven years ago by Dr. H. W. Nichols, is of possible interest in connection with both this impedance and the question of superfluous units. He derives the impedance from the Poynting vector by simple substitutions. Specific use is made, however, of five systems of units. The letter also supplies an illustration of confusion arising from the multiplicity of units in use. Apparently, Heaviside's 30 ohms ("Electrical Papers," ii, p. 377, 1888) was in ordinary ohms and not in Heaviside's own units, as Nichols quite naturally assumed. The correct explanation of the 30 ohms seems to be that Heaviside's "resistance-operator of an infinitely long tube of unit area" was not intended to be the characteristic impedance, as I define it.

In definitive units the characteristic impedance of free space equals the square of the effective volts per meter, in a plane electromagnetic wave, divided by the transmitted watts per square meter. For a numerical example, take the figures for strong sunlight (Maxwell, ii, footnote p. 441) which correspond to 666.1 effective volts per meter and 1176 watts per square meter. The characteristic impedance of free space implicitly assumed was thus 377.3 ohms, which checks well with my 376.54 international ohms.

If free space could be bounded in one direction by a thin, plane film having surface resistivity equal to the characteristic impedance of space, a normally incident plane wave would be completely absorbed by the film; there would be neither reflected wave nor transmitted wave beyond the

film. This picture is suggested by the analogy of a transmission line terminated, at the receiving end, in its characteristic impedance, so that there is no reflected wave. The difficulty with the analogy is that free space exists beyond the film and cannot be cut off. This idealized picture may serve, however, to indicate the simplification made possible by the introduction of characteristic impedances in practical problems involving reflection, refraction and absorption.

The characteristic impedance of free space may be usefully introduced into formulas for the characteristic impedances of transmission lines. Thus, assuming perfect conductors, we have:

For flat strips, width w, separation d, if w/d is large or the guard-ring method is employed in measurements,

$$K_f = 376.54 \frac{d}{w};$$

For concentric cylinders, with radii b and a,

$$K_c = \frac{376.54}{2\pi} \log \frac{b}{a}.$$

These characteristic impedances will each agree with the characteristic impedance of free space if $w = d$ and $b = 535.49\, a$. Since these strips are not wide compared with the separation, it would be necessary to employ the guard-ring method to maintain the plane wave assumed in the square shaft between the two strips. These two characteristic impedances would each become one ohm if $w = 376.54\, d$, and $b = 1.0168\, a$.

Practically, the finite conductivity of copper would add a reactance component and change the resistance component. It would be interesting to investigate simple cases numerically and include mutual characteristic impedances between two metallic circuits.

My own interest in the applications of the impedance concept to the electromagnetic field theory dates back to the last quarter of 1931.

On the Theory of Space Charge Between Parallel Plane Electrodes

By C. E. FAY, A. L. SAMUEL and W. SHOCKLEY

The problem of the potential distribution, current, and electron transit time resulting from the perpendicular injection of electrons into the space between parallel planes is considered. The electrons are assumed to be injected uniformly with velocities corresponding to the potential of the plane through which they are injected. Consideration of all possible solutions of the basic equation shows that four general types of potential distribution are possible. Curves are given which enable the easy calculation of transmitted current and transit time and show the complete potential distribution for any concrete example. The case for current injected through both planes is also considered.

The complete mathematical treatment is given in the appendix.

SPACE charge has been studied extensively since the publication of the first papers on the subject by Child [1] in 1911 and Langmuir [2] in 1913. These papers in common with many which followed [3, 4, 5, 6, 7] dealt for the most part with potential distributions which occur when electrons are injected into a region with relatively small initial velocities.

The problem of space charge between parallel planes when the electrons possess arbitrary initial velocities, although contained implicitly in some of this early work, was first considered in detail by Gill [8] in 1925. Gill appears to have clearly understood the phenomenon but did not publish a complete analysis. Other workers, beginning with Tonks [9] in 1927, have considered various aspects of the problem,[10, 11] and recently Plato, Kleen, and Rothe [12, 13] published an extensive analysis. They did not include transit time calculations and their published curves are not easily adaptable for numerical

[1] C. D. Child, *Phys Rev.*, **32**, 492 (1911).
[2] I. Langmuir, *Phys. Rev.*, **2**, 450–486 (1913); also *Phys. Zeits*, **15**, 348 (1914).
[3] W. Schottky, *Phys. Zeits*, **15**, 526 and 624 (1914).
[4] P. S. Epstein, *Vehr. d.D. Phys. Ges.*, **21**, 85 (1919).
[5] T. C. Fry, *Phys. Rev.*, **17**, 441 (1921); *Phys. Rev.*, **22**, 445 (1923).
[6] I. Langmuir, *Phys. Rev.*, **21**, 419 (1923).
[7] I. Langmuir and K. B. Blodgett, *Phys. Rev.*, **22**, 347 (1923); *Phys. Rev.*, **24**, 49 (1924).
[8] E. W. B. Gill, *Phil. Mag.*, **49**, 993 (1925); *Phil. Mag.*, **10**, 134 (1930).
[9] L. Tonks, *Phys. Rev.*, **30**, 501 (1927).
[10] H. C. Calpine, *Wireless Engineer*, **13**, 473 (1936).
[11] Myers, Hartree and Porter, *Proc. Roy. Soc.* **158**, 23 (Jan. 1937).
[12] G. Plato, W. Keen, and H. Rothe, *Zeit. f. Phys.*, **101**, 509 (July 1936).
[13] H. Rothe and W. Keen, *Telefunken Rohre* No. 9 (April 1937).

calculations. In this country Salzberg and Haeff [14] have presented a solution, apparently quite complete, which has not as yet been published. The present independently derived solution differs from other published work in the manner of presentation and in an attempt at completeness. The results are presented in the form of curves which give potential distributions, electron transit times, and current-voltage relations, directly in terms of units which have a simple physical significance.

Statement of the Problem

In common with some of the earlier treatments, the present discussion will be restricted to a consideration of the steady state potential distributions which can exist in an evacuated region between two parallel planes at known potentials when electrons having normal velocities corresponding to these potentials are injected into the region through one or both planes. For definiteness the case in which all of the current is injected through one plane will be considered first. It will then be shown that the results may be applied to the more general problem. It is supposed that the electrostatic potential is a function of one rectilinear coordinate only and that all the electrons move parallel to this coordinate and have the same total energy. The assumption of equal energy electrons leads to indeterminancy at planes where the potential and its gradient are both zero. At these planes the existence of non-uniform electron velocities will be recognized in so far as it provides a selective mechanism to resolve the mathematical indeterminancy.[15]

Units

It has been found convenient to express voltages, currents, distances and transit times in terms of some derived units which are related to these quantities in certain ways by the simple Child's law equation for space charge limited current and by the corresponding transit time equation. The use of such derived units makes it possible to present a limited number of curves which are then applicable to a wide variety of conditions and may be used with a minimum amount of computations. The physical significance of the units also simplifies the interpretation of the results.

In order to understand how these units are obtained consider the hypothetical situation shown in Fig. 1 in which an electron current from a space charge limited cathode is injected into the region to the

[14] B. Salzberg and A. V. Haeff, *Proc. I.R.E.*, **25**, 546 (May 1937).—Abstract only. This paper appears in full in the January 1938 issue of the *R.C.A. Review.*

[15] This procedure has been justified by Langmuir. See *Reviews of Modern Physics*, vol. 3, pp. 237–244 (1931).

right of a plane at a potential V_1.[16] From Child's equation the current in amperes per square centimeter is given by

$$I = 2.33 \times 10^{-6} \frac{V_1^{3/2}}{S_0^2} = \frac{a^2 V_1^{3/2}}{S_0^2} \text{ (amps. per sq. cm.)} \qquad (1)$$

or solving for S_0

$$S_0 = 1.527 \times 10^{-3} \frac{V_1^{3/4}}{I^{1/2}} = \frac{a V_1^{3/4}}{I^{1/2}} \text{ (centimeters).} \qquad (2)$$

Accordingly, whenever the conditions at the first plane are represented

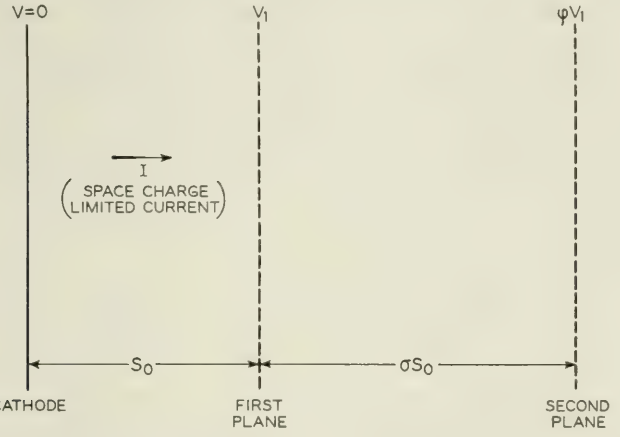


Fig. 1—Hypothetical conditions to assist in visualizing the unit of distance S_0.

by I and V_1, distance from the first plane may be measured in units of S_0. The distance in centimeters is then

$$S = \sigma S_0. \qquad (3)$$

Similarly a natural unit of potential is V_1 and the potential of any plane at a distance S is then given by

$$V = \varphi V_1. \qquad (4)$$

Potential Distributions

All possible potential distributions to the right of the first plane can now be expressed in terms of φ as a function of σ. The mathematical derivation is straightforward and is given in the appendix.

[16] This analogy is useful in getting a physical picture of the units but it must not be carried too far as will be evident when the problem of reflected currents is considered.

The results are shown in Figs. 2, 3 and 4. Figures 6 to 11 refer to current voltage relationships discussed in a later section.

To find the potential distribution when a second plane at a distance σ away from the first plane is held at a potential φ, one enters the figures with the values of φ and σ. Any member of the family of curves drawn in heavy solid lines passing through this point represents a possible potential distribution. The regions occupied by the heavy curves are bounded by certain limiting curves identified by lower case letters. These curves correspond to certain space charge conditions between the planes and are indicated by the same letter whenever met. The dashed curves refer to transit time and are explained below. Situations corresponding to low potentials and large spacings occur, for the most part, on Fig. 2, while those for high potentials and small spacings occur on Fig. 3. An overlap region exists for which potential distributions may be found on both families of curves. The interpretation of this is found by inquiring more closely into the nature of the curves of Figs. 2, 3 and 4.

There are in all four distinct types of potential distributions. These are:

Type *A*—The second plane is at a negative potential. All the injected current is reflected. Potential distributions are the same as for the case of temperature limited emission with the current $2I$ from the zero potential plane.

Type B—Both planes positive with potential zero between them. Injected current partially transmitted and partially reflected. Potential distributions corresponding to space charge limited emission on both sides of a virtual cathode at the zero of potential.

Type C—Both planes positive with a potential minimum (at a positive potential) between them. Complete transmission of injected current.

Type D—Both planes positive with no minimum between them. Complete transmission of injected current.

The curves on Fig. 2 are seen to fall into two groups, those which extend to negative values of φ (marked with values of β) and those which contain the value $\varphi = 0$ but which remain positive. The first group (Type A) obviously corresponds to conditions under which all of the injected current returns to the first plane. The slopes of these potential curves at the reflection planes are not zero but are related to the parameter β (shown on the curves) by the equation

$$\frac{d\varphi}{d\sigma} = -\frac{4}{3}\sqrt{2}\beta^{1/4}. \tag{5}$$

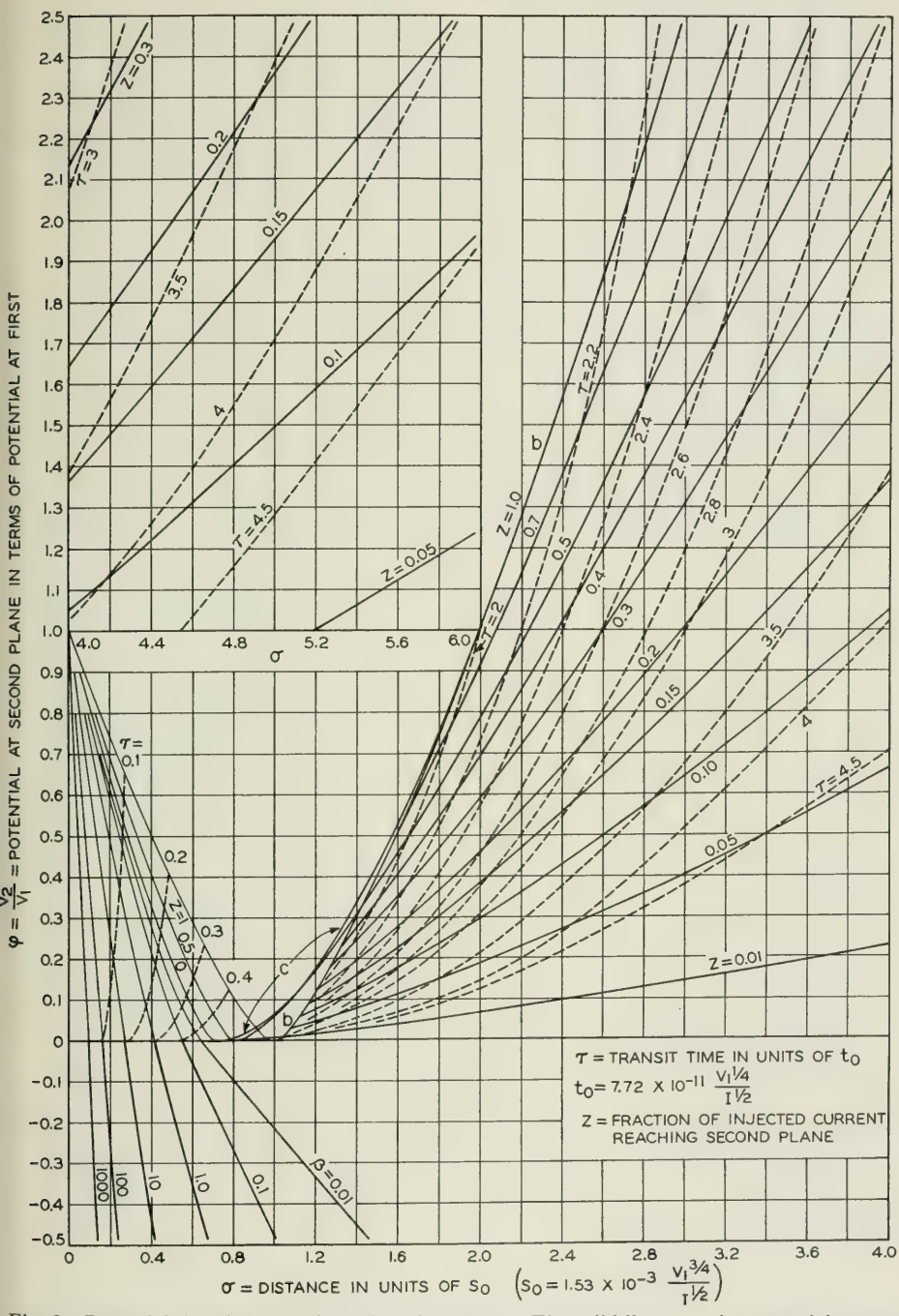


Fig. 2—Potential distributions of the A and B types. The solid lines are the potential curves; the broken lines indicate the transit times.

For the second group of curves (Type B) a certain fraction of the current denoted by Z (and so indicated on the curves) will be transmitted while the fraction $1 - Z$ will be reflected toward the first plane. These potential curves have zero slopes at the reflection planes so that they correspond to solutions of Child's equation on both sides of a so-called virtual cathode existing at the reflection plane. For each value of Z from 0 to 1 a possible potential distribution is obtained.

It will be observed that the portions of the solid curves to the left of the zero points are drawn lighter than the rest. These portions correspond to potential distributions resulting from conditions with reflected current and so while applying as extensions of the heavy curves to the right they cannot be entered directly with values of φ and σ. Should values used to enter the figure fall in this region the absence of a potential zero is indicated and the correct distribution is obtained by entering Fig. 3. Before leaving Fig. 2 the existence should be noted of a small domain lying between the curves marked b and c for which two different B solutions are possible for the same injected current, spacing, and potential, one corresponding to larger values of Z than the other. The significance of this region will be apparent when the current voltage relationships are considered. All solutions of the A and B types are represented on Fig. 2.

Solutions of the C and D types are to be found on Fig. 3 except for a small overlap region which for clearness is shown on Fig. 4. Values of φ and σ which cannot be entered on Fig. 2 as well as values common to both B and C types are to be found on these figures. The additional overlap solutions of the C type are shown in Fig. 4. These are extensions of the potential distribution curves of Fig. 3 after they reach the right-hand boundary curve where they turn inward as shown and overlap the other solutions.

For convenience each curve of the C type is labeled by the value of its minimum potential. The parameters of the D curves do not have this simple physical significance. However, for all cases the value of the curve parameter is simply related to the electric field at the first plane by the equations:

type C
$$\frac{d\varphi}{d\sigma} = -\frac{4}{3}\sqrt{1 - \varphi_{min.}}^{1/2}. \tag{6}$$

type D
$$\begin{cases} \dfrac{d\varphi}{d\sigma} = \dfrac{4}{3}\sqrt{1 - \alpha}^{1/2}, & (7) \\ \dfrac{d\varphi}{d\sigma} = \pm\dfrac{4}{3}\sqrt{1 + \beta}^{1/2}. & (8) \end{cases}$$

type A
$$\frac{d\varphi}{d\sigma} = -\frac{4\sqrt{2}}{3}\sqrt{1 + \beta}^{1/2}. \tag{9}$$

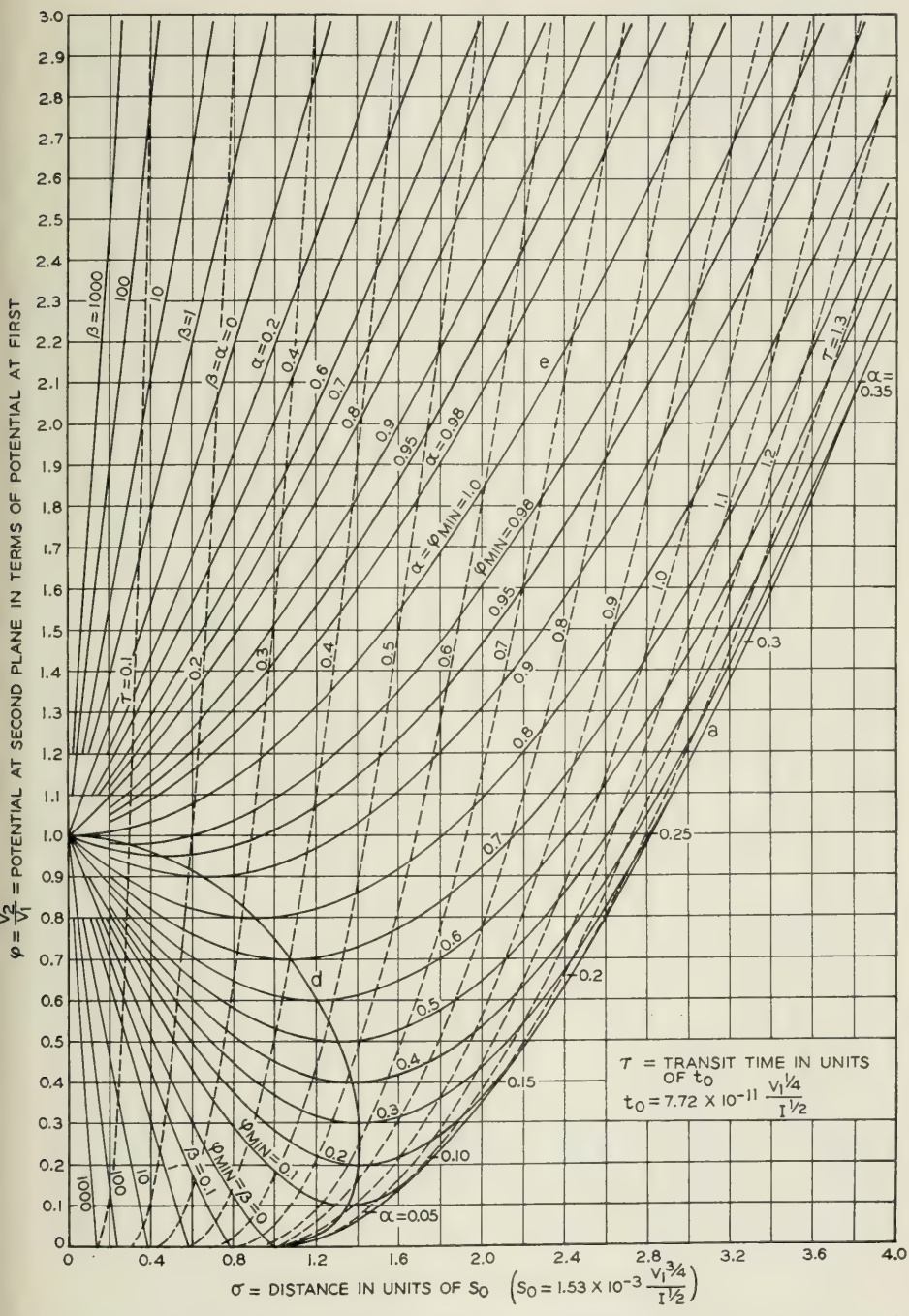


Fig. 3—Potential distributions of the C and D types. The solid lines are potential curves, the broken lines indicate the transit times.

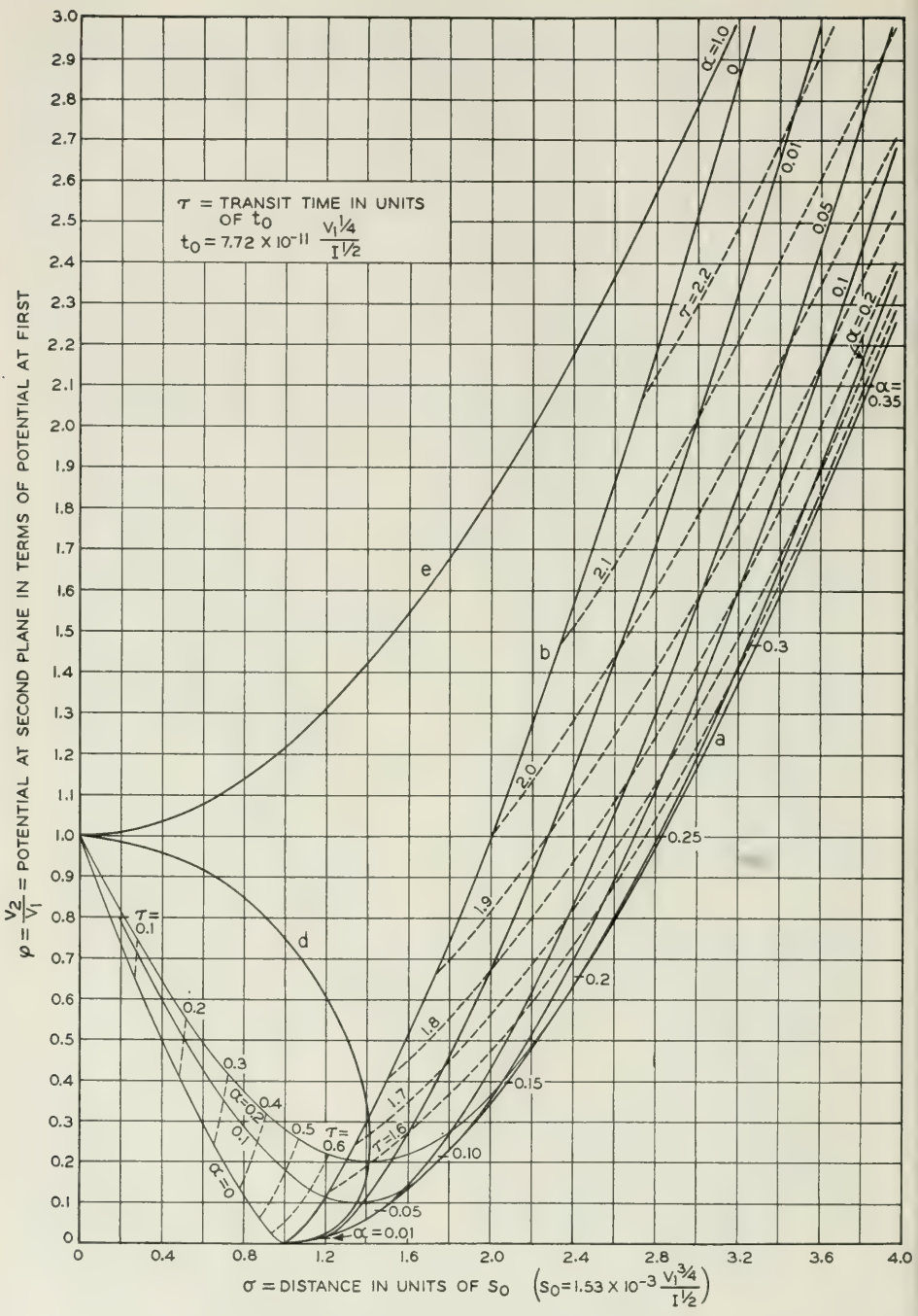


Fig. 4—Potential distributions of the C overlap type.

It should be noted that while some of the curves in the D region shown in Fig. 3 resemble the portions of the curves drawn in light lines on Fig. 2, the curves in Fig. 3 have essentially different physical content, representing solutions in which all of the injected current is transmitted. Values of σ and φ may, therefore, be entered in all regions on Fig. 3 while entering values on Fig. 2 are restricted to the regions covered by the heavy lines.

The occurrence of overlap regions indicates that for certain boundary conditions more than one type of potential distribution may exist. In the practical case the choice between these various possible solutions depends upon the manner in which the boundary conditions are established.[17]

Current Injection from Both Sides

The present analysis although derived on the assumption of current injection from one side only is equally applicable to the situation shown schematically in Fig. 5. The potential distribution curves

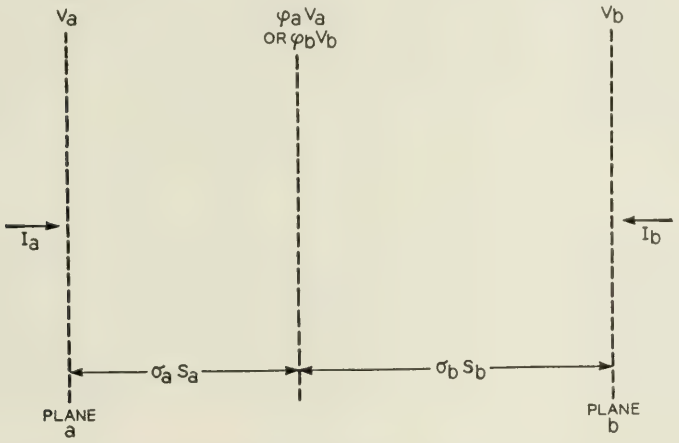


Fig. 5—Schematic representation of the general conditions under which the potential distribution analysis may be applied.

which occur when the currents I_a and I_b are injected from opposite sides will be identical with those obtained for an assumed injection from one side only of a current I equal to the numerical sum of I_a and I_b. Values of V_1, φ and σ are chosen to correspond with an assumed direction of injection and these values are entered on the figures. For solutions of the B type, different distributions will result depending

[17] This matter will be treated in more detail in the section on circuit characteristics.

upon the assumed direction of I. In some instances all solutions may correspond to physically realizable distributions for the double injection case. The solutions will, in general, require that the potential zero be located at different places and may require that the net flow of current across the zero plane be in opposite directions. The physical reality of each of these distributions must be checked by noting the value of Z for the indicated solution and solving for the derived current $(2 - Z)I$ on the injected side of the zero potential plane and the derived transmitted current ZI. If these are respectively greater than the numerical values of I_a and I_b (paired consistent with the assumed direction of injection) the indicated solution is possible. In addition there may exist a solution in which the currents I_a and I_b are each totally reflected at two different zero potential planes, separated by a region of zero potential. The possible existence of this solution must be tested for separately.

Electron Transit Time

So far no use has been made of the dashed curves shown in Figs. 2, 3 and 4. These give the electron transit time from the first plane to any desired plane. The significance of the unit of time is to be found by again referring to Fig. I. The time an electron takes to travel from the cathode to the first plane is given by

$$t_0 = c\frac{V_1^{1/4}}{I^{1/2}} = 7.72 \times 10^{-11}\frac{V_1^{1/4}}{I^{1/2}} \text{ (seconds)}. \tag{10}$$

This value of t_0 is a natural unit of time to be used whenever the conditions at the first plane are expressed by V_1 and I just as S_0 is a natural unit of distance. Accordingly, electron transit times from this plane to any other plane are expressed in units of t_0. The time in seconds is then

$$T = \tau t_0. \tag{11}$$

Transit times for reflected electrons are not given but may be computed by taking the time to the reflection plane and adding to it the returning time. This latter is obtained by taking the difference between the time to the reflection plane and the direct time to the plane under consideration.

Circuit Characteristics

In many practical cases the first plane will coincide with a mesh or grid electrode in a vacuum tube and the second plane will correspond with a plate electrode. Under these conditions the current-voltage

relations which exist at the second plane are of considerable interest. Since the distance between the planes is now fixed, the parameter σ is not appropriate. However, the spacing d between planes and the voltage V_1 serve to define a natural unit of current density i_0, which is indicated by the hypothetical situation of Fig. 6. The value of i_0 is

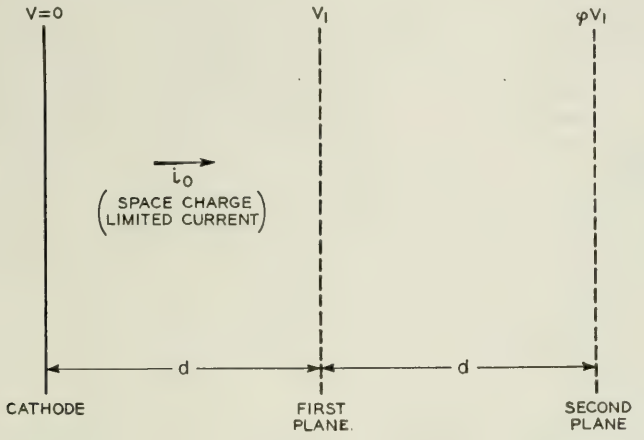


Fig. 6—Hypothetical conditions to illustrate the significance of the unit of current i_0.

$$i_0 = a^2 \frac{V_0^{3/2}}{d^2} = 2.33 \times 10^{-6} \frac{V_1^{3/2}}{d^2} \text{ (amperes per sq. cm.)} \qquad (12)$$

The current in units of i_0 is denoted by γ so that the injected current I in amperes per sq. cm. is given by

$$I = \gamma i_0. \qquad (13)$$

Similarly, the transmitted current under conditions corresponding to type B potential distributions will be

$$ZI = Z\gamma i_0. \qquad (14)$$

Specifying γ and φ is equivalent to specifying σ and φ.[18] For this reason many of the limiting curves of the circuit characteristic plots are simply related to those shown in the potential distribution plots and are designated by the same letters. The regions for which the different types of potential distributions may exist are shown in Fig. 7 now in terms of γ and φ. In Fig. 8 the regions are defined in terms of φ and the transmitted current $Z\gamma$. The boundaries of the C and D

[18] This is a consequence of the relationship $\gamma = \sigma^2$, which is derived in the appendix.

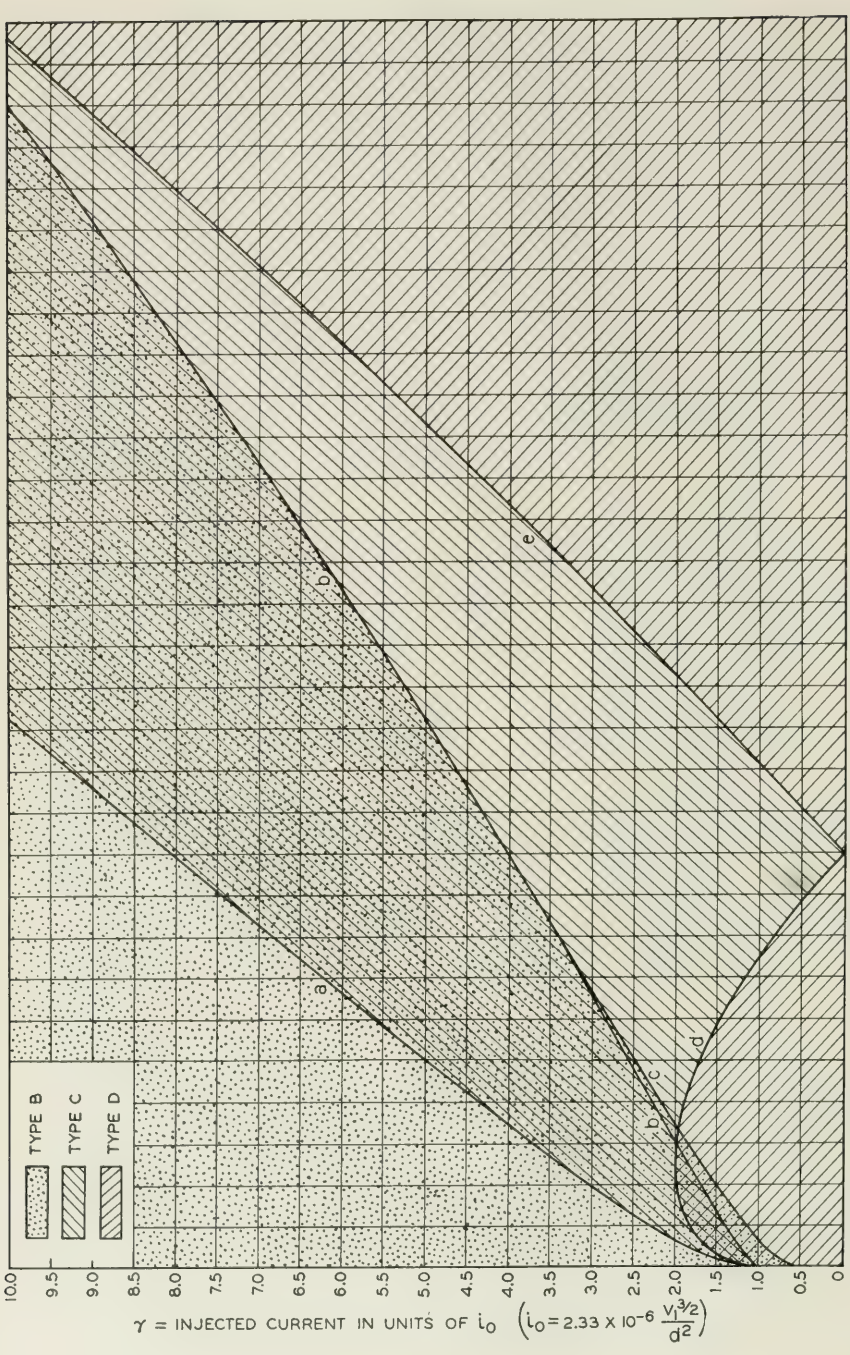


Fig. 7—Key chart giving boundary conditions for the different distributions in terms of the potential and the injected current.

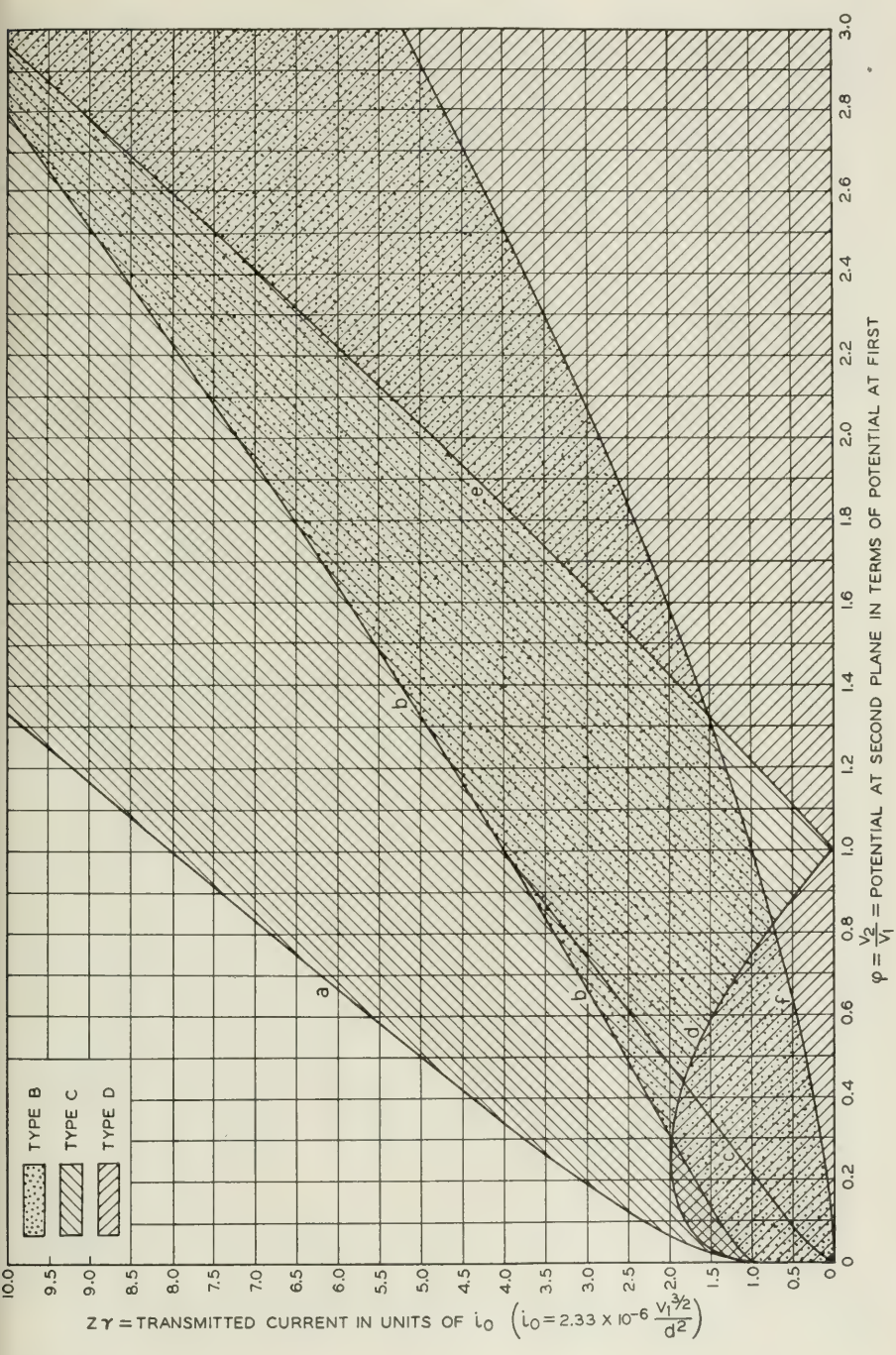


Fig. 8—Key chart giving boundary conditions for the different distributions in terms of the potential and the transmitted current.

regions are of course the same in the two plots (i.e., $Z = 1$) while the boundaries of the B region are changed, since only the portion $Z\gamma$ of the injected current is transmitted. The A region, which corresponds to negative values of φ for all values of γ and gives $Z\gamma = 0$, is not indicated. It may be of interest to note that for any given value of φ it is impossible to transmit more than a certain maximum current, shown by curve a, and that this maximum occurs for a C type distribution. For injected currents greater than this amount only type B distributions are possible.

An enlargment of the B region is shown in Fig. 9a and a portion of the region to a still larger scale in Fig. 9b. Lines for constant injected current γ are plotted with the potential φ and the transmitted current $Z\gamma$ as the coordinates. These plots correspond to plots frequently used to describe vacuum tube characteristics and may be used accordingly. For values of γ less than four and values of φ less than unity, double values of $Z\gamma$ appear. In the region lying between the lines b and c on Fig. 9, the slopes of the constant γ lines are negative corresponding to conditions which are unstable unless sufficient resistance is included in the external circuit. Conditions outside of this region are always stable and need no further comment.

Solutions of the C type are shown in Fig. 10a. For these solutions the injected and transmitted currents are equal. The depth of the potential minimum between the two planes is, however, of interest and is shown in the figure in terms of the lesser of the two boundary potentials. For values of φ greater than unity the value of φ min. (i.e., the minimum potential in units of V_1) is indicated while for values of φ less than unity the value of $\varphi_{min.}$ (i.e., the minimum potential in units of V_2) is shown. In the enlargement, Fig. 10b, both values are given. For values of $\varphi_{min.} = 1$, curve e, and $\varphi_{min.} = 1$, curve d, the potential minimum is equal to the potential at one of the planes and is located at this plane. For currents greater than the value indicated by these limiting curves the potential minimum becomes deeper and moves away from the plane while for lesser currents it disappears entirely. The minimum value to which the potential may be forced by increasing the injected current before the distribution changes precipitously to one of the B type is indicated by the values on curve a.

Again to avoid confusion the overlap type C region is shown separately in Fig. 11. Conditions are somewhat different for this region in that with any assumed value of φ a minimum potential of 0 (but with complete transmission of injected current) occurs for a certain value of γ. With increasing injected current, the potential minimum

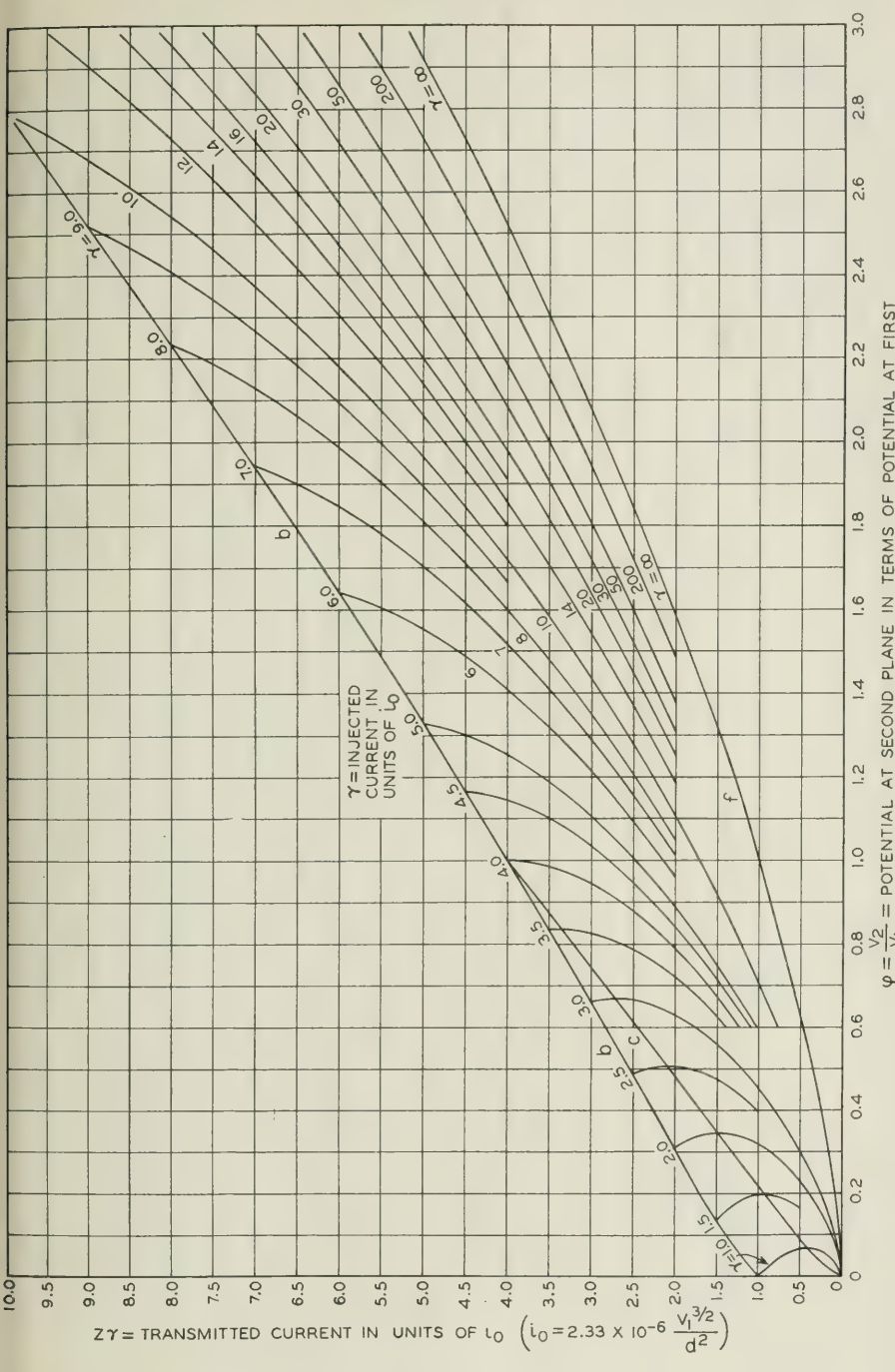


Fig. 9*a*—Electrical characteristics for type B distributions.

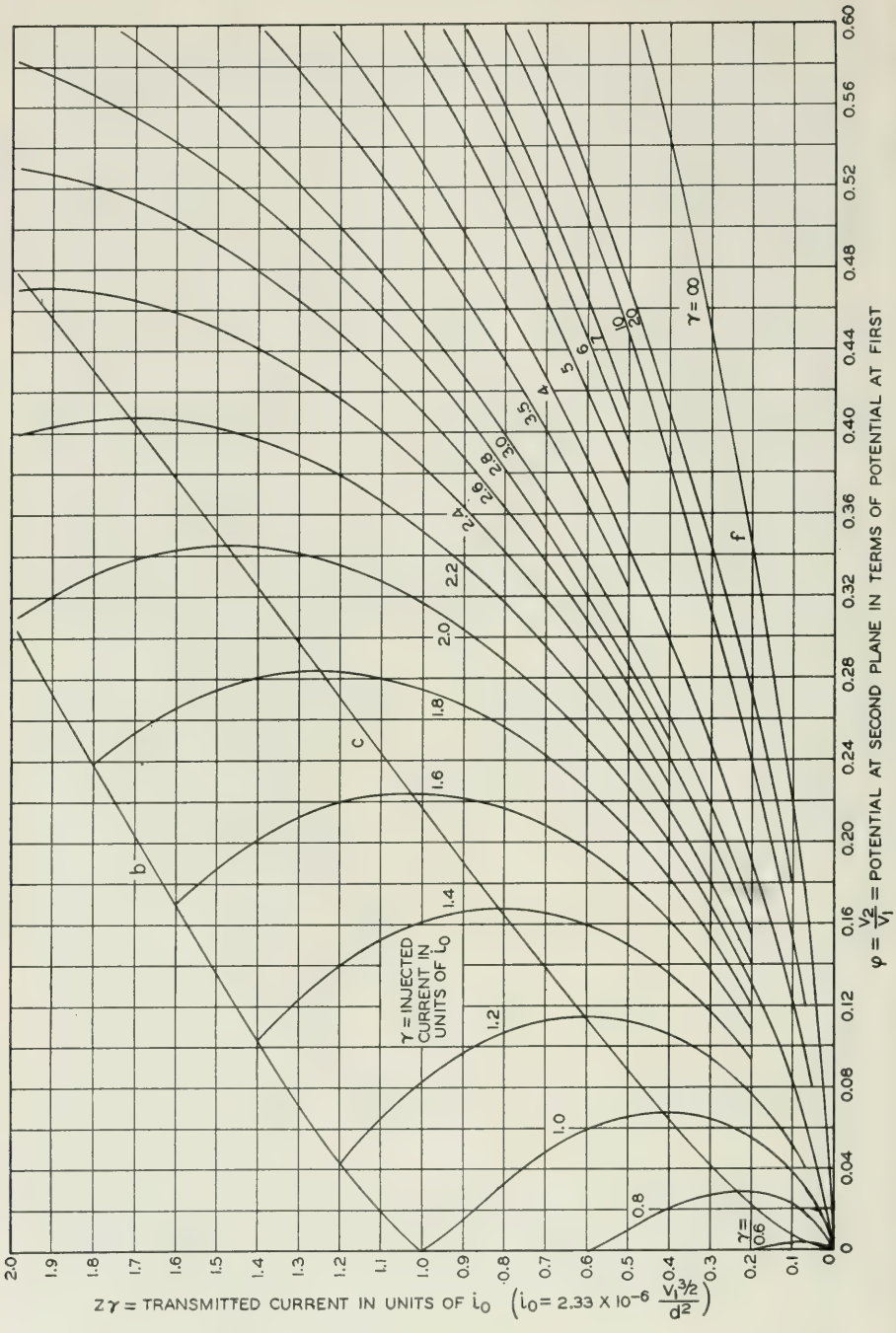

Zγ = TRANSMITTED CURRENT IN UNITS OF i_0 $\left(i_0 = 2.33 \times 10^{-6} \frac{V_1^{3/2}}{d^2}\right)$
φ = $\frac{V_2}{V_1}$ = POTENTIAL AT SECOND PLANE IN TERMS OF POTENTIAL AT FIRST
γ = INJECTED CURRENT IN UNITS OF i_0
γ = ∞
γ = 0.6
b
c
f

Fig. 9*b*—Enlargement of a portion of Fig. 9*a*.

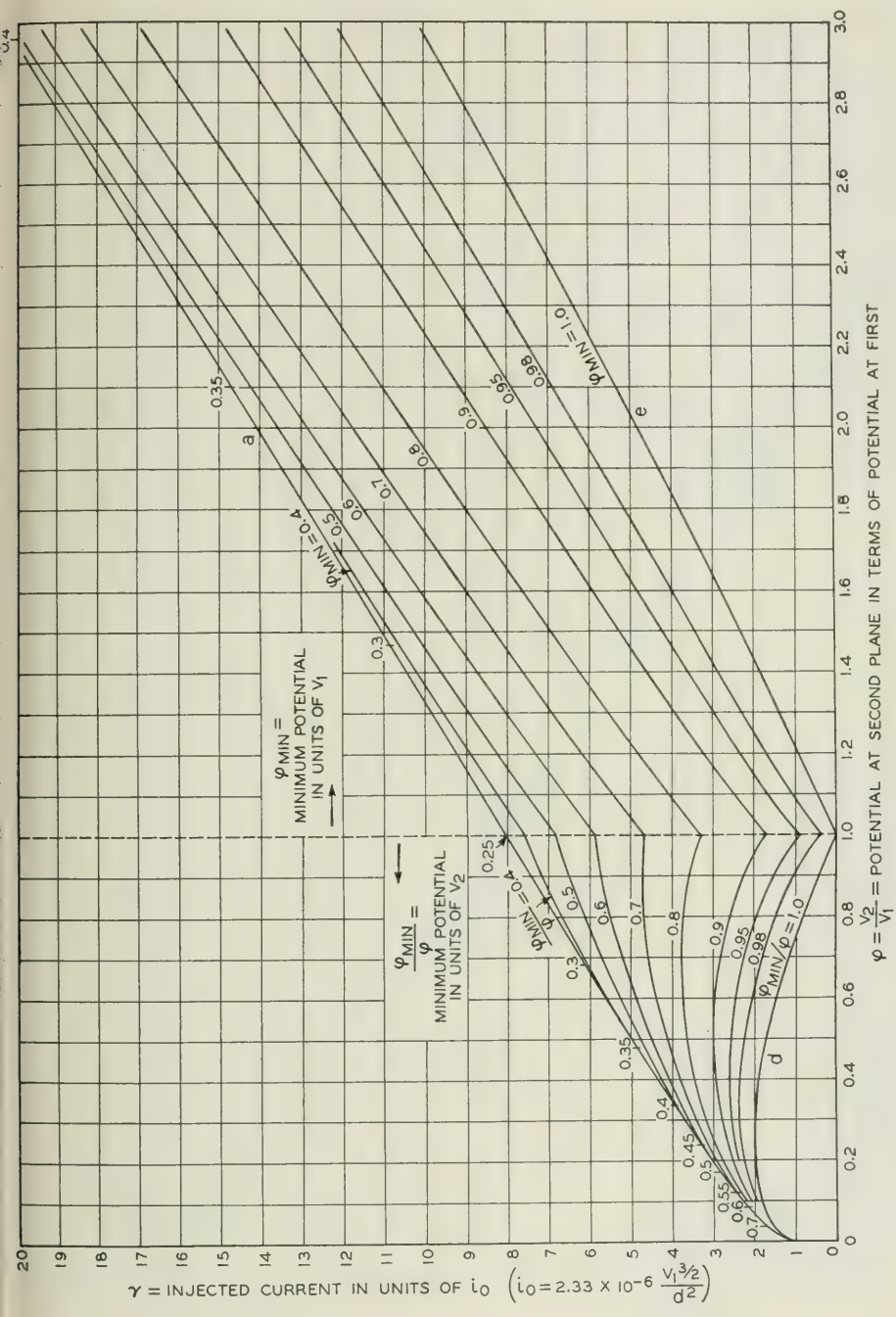

γ = INJECTED CURRENT IN UNITS OF i_0 $\left(i_0 = 2.33 \times 10^{-6} \frac{V_1^{3/2}}{d^2}\right)$
φ = $\frac{V_2}{V_1}$ = POTENTIAL AT SECOND PLANE IN TERMS OF POTENTIAL AT FIRST
$\frac{\varphi_{MIN}}{\varphi}$ = MINIMUM POTENTIAL IN UNITS OF V_2
φ_{MIN} = MINIMUM POTENTIAL IN UNITS OF V_1
a
b
c
d
e
φ_{MIN} = 0.4
φ_{MIN} = 0.5
φ_{MIN} = 0.98
φ_{MIN} = 1.0
φ_{MIN}/φ = 1.0

Fig. 10*a*—Electrical characteristics for type C distributions (overlap solutions on Fig. 11).

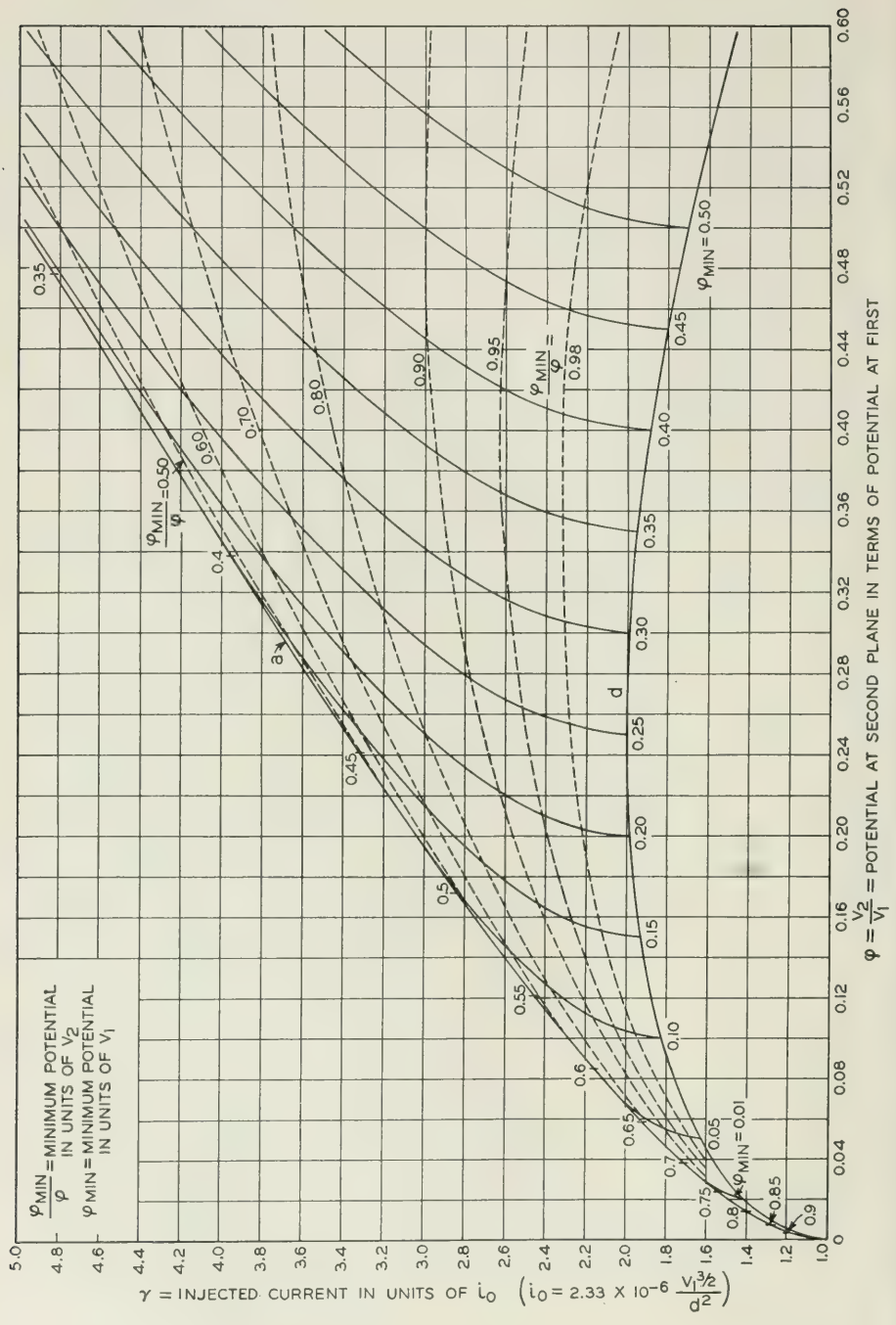

$\frac{\varphi_{MIN}}{\varphi}$ = MINIMUM POTENTIAL IN UNITS OF V_2
φ_{MIN} = MINIMUM POTENTIAL IN UNITS OF V_1
γ = INJECTED CURRENT IN UNITS OF i_0 $\left(i_0 = 2.33 \times 10^{-6} \frac{V_1^{3/2}}{d^2}\right)$
$\varphi = \frac{V_2}{V_1}$ = POTENTIAL AT SECOND PLANE IN TERMS OF POTENTIAL AT FIRST
$\varphi_{MIN} = 0.50$
$\varphi_{MIN} = 0.01$
$\frac{\varphi_{MIN}}{\varphi} = 0.98$
$\frac{\varphi_{MIN}}{\varphi} = 0.50$
a
d

Fig. 10*b*—Enlargement of a portion of Fig. 10*a*.

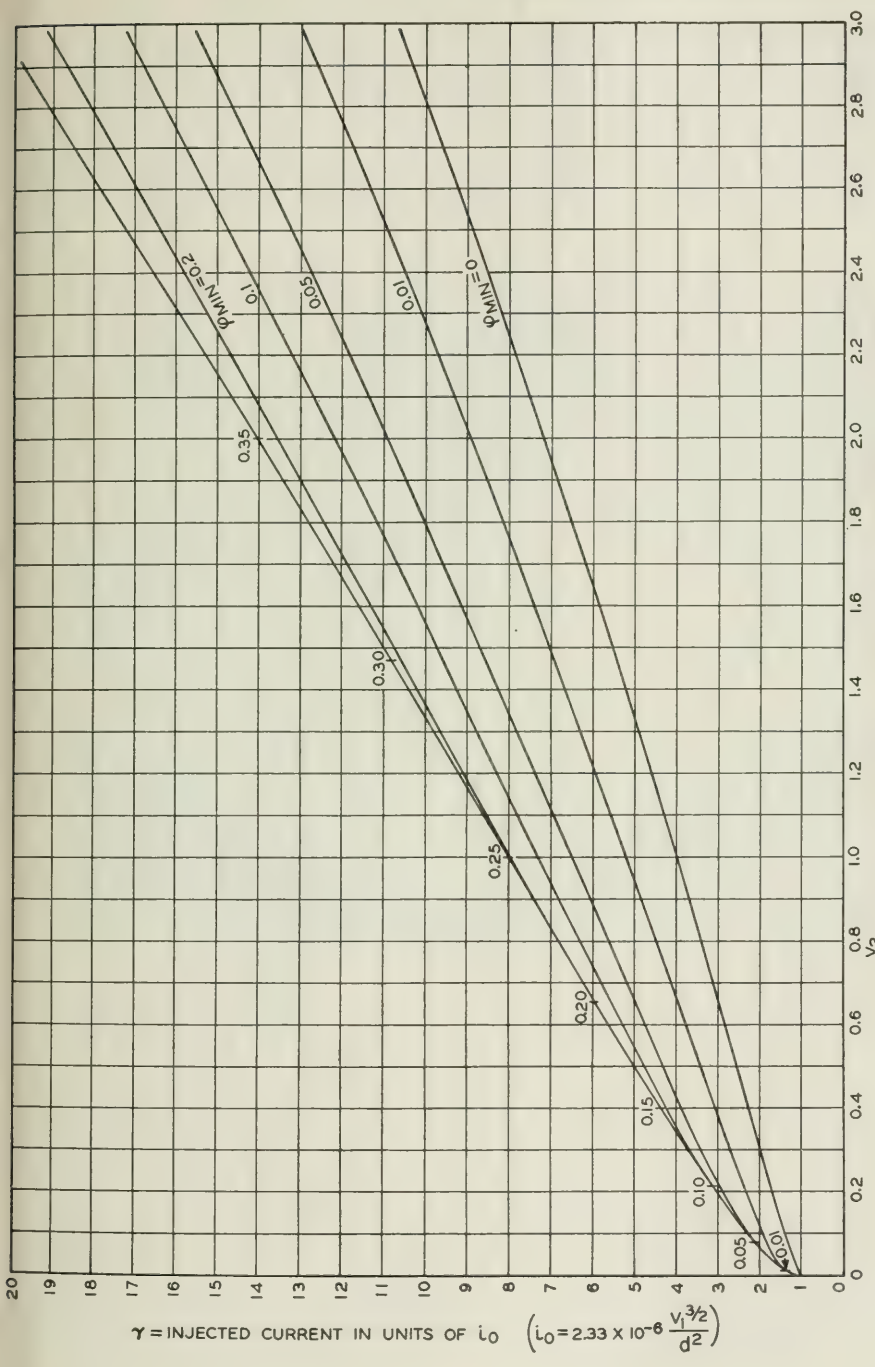


Fig. 11—Electrical characteristics for type C overlap distributions.

rises, finally reaching the limiting values indicated on the upper curve of Fig. 11; the limiting conditions there are precisely the same as for curve *a* of Fig. 10. Further increase of injected current produces a transition to type *B* solutions.

It seems questionable that this overlap type C condition can exist in a practical case. An investigation of this question involving external circuit considerations is beyond the scope of this paper.

Transition Between Distribution Types

The physical choice between the different possible potential distributions which may exist with a given set of boundary conditions is determined by the sequence in which the boundary conditions are established. Extreme values of any parameter are seen from Figs. 7 and 8 to lie in regions for which only one solution is possible. If the boundary conditions are varied slowly and continuously from these values, the indicated type of distribution will persist until the limit of this region is reached at which time a sudden transition must occur to another indicated type of distribution. Inspection of Figs. 7 and 8 will show that at such transitions only one other type of distribution is ever possible. The determination of the correct physical distribution can thus be made without ambiguity.

Certain peculiarities are, however, to be noted. A survey of all possible transitions in which γ and φ are treated as independent variables will indicate that, starting from extreme conditions and changing conditions continuously in the same direction, distributions of the overlap C type shown in Figs. 4 and 11 never occur. A second peculiarity has to do with the unstable region of type B solutions shown on Fig. 9. When this region is entered with insufficient resistance in the external circuit, instability results with a sudden transition to a corresponding stable type of distribution.

The space model shown in Fig. 12 has been found to be of value in visualizing problems involving transitions. The three coordinates used in its construction are the second electrode potential φ (to the right) the injected current γ (to the left) and the transmitted current $Z\gamma$ (vertical). Solutions corresponding to potential distributions of the C and D type, for which $Z\gamma = \gamma$, appear as a celluloid plane inclined at 45 degrees, on which the values of the potential minima are indicated. Solutions of the C overlap type have been omitted. Solutions of the B type are represented by the concave surface of the model. Viewing the model from a different angle as shown in Fig. 13 (where the celluloid sheet is removed) the unstable B region appears as an over-hanging cliff extending between γ values of 0.5 and 4. Bounding curve *c*

separating the unstable region from the stable region is drawn on the model and lies along the vertical surface of the cliff. The small region lying between γ values of 0 and 0.5, for which direct transitions between *A* solutions and *D* solutions occur, can be seen in both views as a small vertical cliff.

Solutions corresponding to varying conditions will be given by the position of a movable particle confined to the surface of the model.

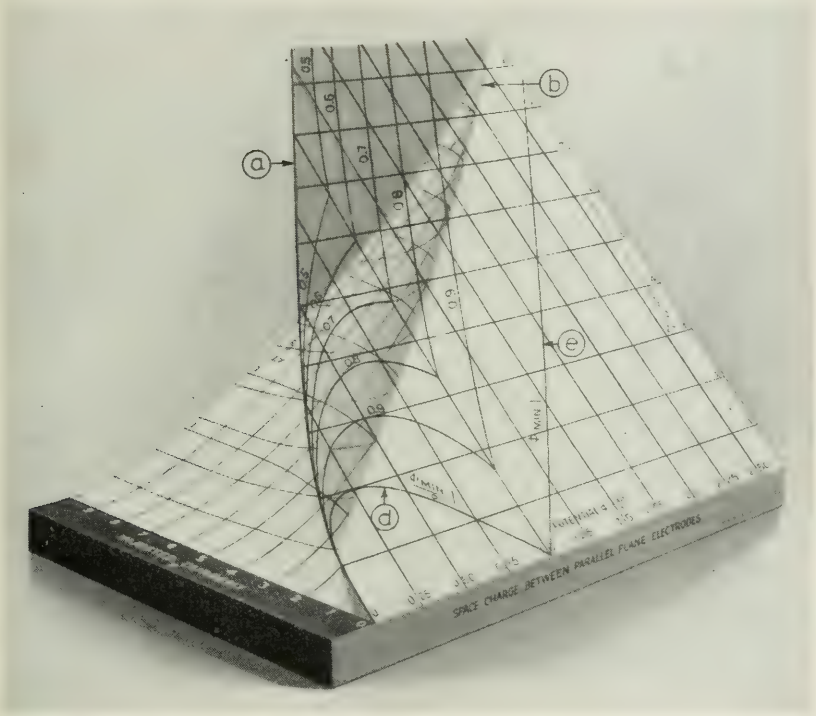


Fig. 12—A three-dimensional model illustrating the current-voltage relationships. Solutions of types C and D appear on the celluloid surface and solutions of the B type appear on the concave surface.

When this particle lies on the celluloid surface it may be moved around at will and will not "fall off" unless an attempt is made to go to larger values of the injected current γ than is permitted by the bound-curve *a* (the curved edge of the celluloid). The particle may, however, "fall off" at this edge and land on the concave surface from which it can escape only by climbing up the surface to the bounding curve *b* which is common to both surfaces. The overhanging cliff is, however, particularly treacherous, for unless the particle is supported by the

existence of sufficient resistance in the circuit connected to the second electrode, it will go through the model coming to rest on the celluloid surface above. With no resistance in the external circuit this curious behavior will occur at the vertical part of the cliff indicated by the bounding curve *c*. The presence of some resistance will cause this to occur for larger values of $Z\gamma$. The exact place can of course be determined by noting the point of tangency between a resistance line drawn on Fig. 9 and one of the characteristic curves—just as the stable operating conditions for any negative resistance device is found.

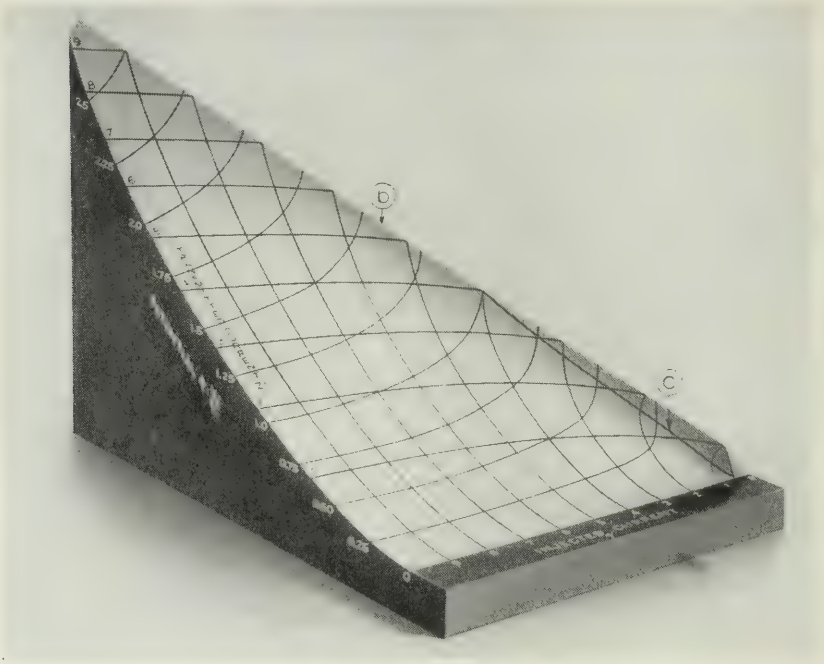


Fig. 13—A different view of the concave surface of the model shown in Fig. 12. (The celluloid surface has been removed.)

The entire surface is stable if the external resistance is greater than approximately 0.25 in units of $\varphi/Z\gamma$.

Illustrative Examples

I. Class B Operation of a "Critical Distance" Tetrode

Since the purpose of "critical distance" operation is to prevent the passage of secondary electrons from the plate (our second plane) to the screen grid (our first plane), the existence of a potential minimum

is desired; however, type B solutions with some reflected electrons would be objectionable. The permissible operating range must therefore lie in the C region on Fig. 7. If, for example, we operate the tube with $V_2 = V_1$ as the quiescent point and with a resistance load, the operating curve is a straight line through the point $\gamma = 0$, $\varphi = 1$. The upper excursion of this line is set by the upper limit, curve a, of the C region. For characteristics which cross curve d and enter the D region a loss of suppression would result. Maximum efficiency occurs with the largest possible excursion of V_2, that is for a line tangent to the d curve at the quiescent point. The minimum instantaneous value of V_2 for the conditions specified is seen to be 25 per cent of its quiescent value as contrasted with a usual value of about 10 per cent for a pentode when similarly operated. For the instantaneous value of V_2 to fall to 10 per cent of its initial value (without loss of suppression) the quiescent point must be at $\gamma = 0$, $\varphi = 1.3$ with the operating line tangent to curve d at about $\varphi = 0.5$. Operation in the D region above $\varphi = 1$ does not result in a secondary current from plate to grid.

II. Triode Operation in the Positive Grid Region

Figure 1 may be taken as representing an idealized triode. The second plane represents the plate so that $V_2 = V_p$ and the first plane represents the grid plane with effective potential V_1. We let

$$V_1 = V_g + V_p/\mu,$$

where μ will be a variable if space charge is formed in the plate-grid region and may differ considerably from its "cut-off" value.

If the fraction of the electrons stopped by the grid is negligible and if operation does not enter the B region, with disturbing effects produced by reflected electrons, then Fig. 1 becomes an accurate representation and in view of the choice of units σ is a geometrical constant of the tube. Hence the operating characteristic is a constant γ line ($\gamma = \sigma^2$) on Fig. 7.[19]

For normal operation the characteristic must not extend into the B region. If the grid is to be driven far positive it is desirable to stay in the C region for all values of $\varphi < 1$ to prevent secondary electrons from leaving the plate. The minimum value of φ as a function of γ may be read from the intersections of the constant γ lines with limiting curve a if $\gamma > 2$ and limiting curve d if $\gamma < 2$. Now since

$$\varphi = \frac{V_P}{V_1} = \frac{V_P}{V_g + \dfrac{V_P}{\mu}}$$

[19] Proof of the relationship $\gamma = \sigma^2$ is contained in the appendix.

and

$$V_g/V_P = \frac{1 - \varphi/\mu}{\varphi} = \frac{1}{\varphi} - \frac{1}{\mu},$$

if μ and σ are known, the maximum value of V_g/V_P may be calculated. Fig. 7 shows that a value of γ just greater than 2 will give the lowest minimum value for φ ($\varphi = 0.07$) by intersection with curve a. In this region $\varphi/\mu \ll 1$ so that

$$\left(\frac{V_g}{V_P}\right) \text{max.} = \frac{1}{\varphi} = 14.$$

For values of γ only slightly less than 2 the minimum value of φ rises abruptly, being given by the intersection with curve d, while for values greater than 2 the rise in φ is not so abrupt.

It should be noted that for the usual cylindrical triode the departure from the parallel plane case is not very marked in the grid-plate region while it can be taken into account by the usual cylindrical formulæ for the cathode grid region. It therefore becomes possible to extend this analysis in an approximate sort of way to cover the cylindrical case.

APPENDIX

Definition of Symbols

V_1 = Potential of the first plane in volts,
V = Potential of a second plane in volts,
φ = Ratio of the potential at the second plane to the potential of the first plane,
I = Injected current in amperes per sq. cm.,
x = Distance in centimeters,
d = Spacing between the two planes in centimeters (used when this distance is regarded as a fixed quantity),
S = Distance between the two planes in centimeters (used when this distance is regarded as a variable),
σ = Distance between the planes expressed in units S_0,

$$S_0 = 1.527 \times 10^{-3} \frac{V_1^{3/4}}{I^{1/2}} \text{ (centimeters)},$$

$$a = 1.527 \times 10^{-3} = [10^{13}(2e/mc)^{1/2}/9\pi c^2]^{1/2},^{20}$$

$a^2 = 2.33 \times 10^{-6}$,
Z = Fraction of the injected current I which reaches the second plane,
t = Time in seconds,
T = Electron transit time between the planes in seconds,

τ = Electron transit time expressed in units of t_0,

$t_0 = 7.72 \times 10^{-11} \frac{V_1^{1/4}}{I^{1/2}}$ (seconds),

$c = 7.72 \times 10^{-11} = 3a(m\mathbf{c}/2e)^{1/2}10^{-4}$,[20]

γ = Injected current expressed in units of i_0,

$i_0 = 2.33 \times 10^{-6} \frac{V_1^{3/2}}{d^2} = \frac{a^2 V_1^{3/2}}{d^2}$ (amperes per sq. cm.),

α = Ratio of the potential at a potential minimum when one exists to the potential of the first plane, otherwise merely a convenient constant of integration. (In the text the symbol $\varphi_{min.}$ is used in the cases where the physical significance can be attached.)

β = An integration constant similar to α but associated with an opposite sign.

Relationship between γ and σ

By definition

$$S_0 = \frac{a V_1^{3/4}}{I^{1/2}} = \frac{S}{\sigma} \tag{1}$$

and

$$i_0 = \frac{a^2 V_1^{3/2}}{d^2} = \frac{I}{\gamma}. \tag{2}$$

Solving (1) for σ

$$\sigma = \frac{S I^{1/2}}{a V_1^{3/4}}. \tag{3}$$

Solving (3) for γ

$$\gamma = \frac{d^2 I}{a^2 V_1^{3/2}}. \tag{4}$$

But since $S = d$ for the conditions under which γ is used

$$\gamma = \sigma^2 \tag{5}$$

Derivations

The space charge equation based upon Poisson's equation and the energy equation for an electron is[21]

$$\frac{d^2 V}{dx^2} = \frac{4}{9a^2} I V^{-1/2}. \tag{6}$$

[20] In this work e is the charge on the electron in e.s.u., m its mass in grams, $\mathbf{c}$ is the ratio of electrostatic and electromagnetic units, 1 volt = $(10^8/\mathbf{c})$ e.s.u., 1 ampere = $(\mathbf{c}/10)$ e.s.u.

[21] See, for example, Dow, "Fundamentals of Engineering Electronics," John Wiley, 1937, pg. 100.

This implies that all of the electrons at any plane have the same speed, a condition which must be borne in mind when the present analysis is applied.

Integration of equation (6) yields

$$\left(\frac{dV}{dx}\right)^2 = \frac{16}{9}\frac{I}{a^2}(V^{1/2} + \text{const.}) \tag{7}$$

The choice of zero for this constant leads to type B distributions. A negative constant gives type C distributions, and some of the type D distributions while a positive value gives type A and type D distributions. These will be considered in detail in the sections which follow.

Transit time solutions are obtained by writing the energy equation for an electron in a conservative field which in practical units is

$$\frac{dx}{dt} = \left(\frac{2eV10^8}{mc}\right)^{1/2} = \frac{3a}{c}V^{1/2}. \tag{8}$$

Solving for t and introducing numerical values

$$t = \int \frac{c}{3a}V^{-1/2}\,dx = 1.68 \times 10^{-8}\int V^{-1/2}dx \tag{9}$$

$$= 1.68 \times 10^{-8}\int \left(\frac{dV}{dx}\right)^{-1} V^{-1/2}dV.$$

Specialization of this equation for the various types is carried out below.

Integration Constant Zero—Type B

If the constant in equation (7) is set equal to zero, the next integration gives Child's equation, applicable to the type B distribution when the correct values of current are used, corresponding to conditions before and after the potential zero. Before the potential zero the total current, i.e., the arithmetic sum of injected and reflected currents, is $(2 - Z)I$ so that

$$(2 - Z)I = \frac{a^2V^{3/2}}{x^2}, \tag{10}$$

where V is the potential x centimeters before the zero. Hence measuring distance from the first plane in units of S_0 and expressing potential in units of V_1,

$$\sigma_- = \frac{1}{(2 - Z)^{1/2}} - \frac{\varphi^{3/4}}{(2 - Z)^{1/2}}. \tag{11}$$

Similar analysis for the region beyond the potential zero where

$$ZI = \frac{a^2 V^{3/2}}{x^2}, \tag{12}$$

yields

$$\sigma_+ = \frac{1}{(2 - Z)^{1/2}} + \frac{\varphi^{3/4}}{Z^{1/2}}. \tag{13}$$

Introducing the identity $\gamma = \sigma^2$ in equation (13) gives the relationship

$$Z\gamma = \left[\left(\frac{Z}{2 - Z} \right)^{1/2} + \varphi^{3/4} \right]^2. \tag{14}$$

Some of the limiting curves associated with the B solutions are closely related, as is implied by the use of a common letter. Curve c in Fig. 2 corresponds to maximum values of φ for fixed values of σ and hence of $\gamma = \sigma^2$. By inspection this is seen to be the same condition as is implied by curve c in Fig. 9. To find these curves we make φ a maximum in (13) with respect to Z holding σ constant and obtain:

(c)
$$\begin{cases} Z = 2\varphi^{1/2}/(1 + \varphi^{1/2}), & (15) \\ \sigma = (1 + \varphi^{1/2})^{3/2} 2^{-1/2}. & (16) \end{cases}$$

From these the various other forms of curve c are readily found by the relationships $\gamma = \sigma^2$ and $Z\gamma = Z\sigma^2$. The curve b of Fig. 2 corresponds to $Z = 1$ and gives in equation (13)

(b)
$$\sigma = 1 + \varphi^{3/4}. \tag{17}$$

The minimum transmitted current for fixed φ, curve f in Figs. 8 and 9, is seen to correspond to $\gamma \to \infty$; for this the virtual cathode recedes to the first plane and

(f)
$$Z\gamma = \varphi^{3/2}. \tag{18}$$

Introducing dV/dx obtained from equation (7) with the correct current and a zero constant into equation (9) and integrating gives

$$T = \frac{cV^{1/4}}{(2 - Z)^{1/2} I^{1/2}} + \text{const.}, \tag{19}$$

which in units of t_0 measured from the first plane with potentials in units of V_1 gives

$$\tau_- = \frac{1 - \varphi^{1/4}}{(2 - Z)^{1/2}}. \tag{20}$$

Similar analysis to points beyond the potential zero yields

$$\tau_+ = \frac{1}{(2 - Z)^{1/2}} + \frac{\varphi^{1/4}}{Z^{1/2}}. \tag{21}$$

Integration Constant Negative—Types C and D

Introducing for the constant in equation (7) a negative value, say $-(\alpha V_1)^{1/2}$, will give a positive value of V (equal to αV_1) for $dV/dx = 0$ with $d^2V/dx^2 > 0$ and must therefore lead to solutions of the C type.

Integrating once more and introducing the unit S_0 gives

$$x = \pm S_0(\varphi^{1/2} + 2\alpha^{1/2})\sqrt{\varphi^{1/2} - \alpha^{1/2}} + \text{const.} \tag{22}$$

Expressing distance from the first plane in units of S_0, we find two possibilities:

$$\sigma_D = +(\varphi^{1/2} + 2\alpha^{1/2})\sqrt{\varphi^{1/2} - \alpha^{1/2}} - (1 + 2\alpha^{1/2})\sqrt{1 - \alpha^{1/2}} \tag{23}$$

and

$$\sigma_- = -(\varphi^{1/2} + 2\alpha^{1/2})\sqrt{\varphi^{1/2} - \alpha^{1/2}} + (1 + 2\alpha^{1/2})\sqrt{1 - \alpha^{1/2}}. \tag{24}$$

The first of these solutions gives a potential distribution rising continuously as σ increases from zero, hence of type D as was anticipated by the subscript. The second solution decreases to a minimum at

$$\sigma_{min.} = (1 + 2\alpha^{1/2})\sqrt{1 - \alpha^{1/2}} \tag{25}$$

and then increases, the equation to the right of the minimum being

$$\sigma_+ = (\varphi^{1/2} + 2\alpha^{1/2})\sqrt{\varphi^{1/2} - \alpha^{1/2}} + (1 + 2\alpha^{1/2})\sqrt{1 - \alpha^{1/2}}. \tag{26}$$

Curves given by equations 23, 24 and 26 are drawn in Fig. 3. If values of σ and φ corresponding to conditions on the boundary planes are entered in the figure, a C solution is indicated only if the point falls upon a curve of the σ_+ type. This curve then gives the potential distribution to the right of the minimum; to the left of the minimum the distribution is given by the σ_- curve with the same value of α, which has the interpretation $\alpha = \varphi_{min.}$ for this case. Points entered on the σ_- or σ_D curves will clearly give D type solutions.

Three equations for limits of the C region can easily be written down on the basis of the above equations:

(b) $$\sigma = 1 + \varphi^{3/4}. \tag{27}$$

(d) $$\sigma = (1 + 2\varphi^{1/2})\sqrt{1 - \varphi^{1/2}}. \tag{28}$$

(e) $$\sigma = (\varphi^{1/2} + 2)\sqrt{\varphi^{1/2} - 1}: \tag{29}$$

The curve a is obtained by making σ_+ a maximum with respect to α while holding φ constant. This gives

(a)
$$\begin{cases} \alpha = \varphi_{min.} = \varphi(1 + \varphi^{1/2})^{-2}, & (30) \\ \sigma = (1 + \varphi^{1/2})^{3/2}. & (31) \end{cases}$$

Here φ and σ are coordinates of a point on the a curve and $\alpha = \varphi_{\min.}$ is the parameter value for the potential distribution curve tangent to curve a at that point. The σ_+ curves give type C solutions for values before the tangent point and give C overlap solutions beyond this point.

All the curves described in this section are readily transformed to current voltage plots by the relationships $\gamma = Z\gamma = \sigma^2$.

The transit times for the various curves are found from equations (7) and (9) using the value $-(\alpha V_1)^{1/2}$ for the constant. Integrating and measuring time from the first plane, they are

$$\tau_D = +(\varphi^{1/2} - \alpha^{1/2})^{1/2} - (1 - \alpha^{1/2})^{1/2}. \quad (32)$$

$$\tau_{c-} = -(\varphi^{1/2} - \alpha^{1/2})^{1/2} + (1 - \alpha^{1/2})^{1/2}. \quad (33)$$

$$\tau_{c+} = +(\varphi^{1/2} - \alpha^{1/2})^{1/2} + (1 - \alpha^{1/2})^{1/2}. \quad (34)$$

Integration Constant is Positive—Type D

Type D solutions include those given by equations (23) and (24).

Other solutions are obtained by giving the integration constant of equation (7) a positive value, say $+(\beta V_1)^{1/2}$. Integrating the equation and measuring distances from the first plane in units of S_0 we obtain the two possibilities:

$$\sigma = +(\varphi^{1/2} - 2\beta^{1/2})\sqrt{\varphi^{1/2} + \beta^{1/2}} - (1 - 2\beta^{1/2})\sqrt{1 + \beta^{1/2}}, \quad (35)$$

which applies for $\varphi > 1$, and

$$\sigma = -(\varphi^{1/2} - 2\beta^{1/2})\sqrt{\varphi^{1/2} + \beta^{1/2}} + (1 - 2\beta^{1/2})\sqrt{1 + \beta^{1/2}}, \quad (36)$$

which applies for $\varphi < 1$. Corresponding transit times are:

$$\tau = +(\varphi^{1/2} + \beta^{1/2})^{1/2} - (1 + \beta^{1/2})^{1/2}. \quad (37)$$

$$\tau = -(\varphi^{1/2} + \beta^{1/2})^{1/2} + (1 + \beta^{1/2})^{1/2}. \quad (38)$$

Integration Constant is Positive—Type A

The potential distribution curves of the A type are identical in form with those of the D type, where $\varphi < 1$, which result from the

positive value of the constant in equation (7). They differ in numerical values in that the current I must be replaced by the value $2I$ to allow for the reflected current. The correct equation is then

$$\sigma = [(1 - 2\beta^{1/2})(1 + \beta^{1/2})^{1/2} - (\varphi^{1/2} - 2\beta^{1/2})(\varphi^{1/2} + \beta^{1/2})^{1/2}]2^{-1/2}. \quad (39)$$

The corresponding transit times are given by

$$\tau = [(1 + \beta^{1/2})^{1/2} - (\varphi^{1/2} + \beta^{1/2})^{1/2}]2^{-1/2}. \quad (40)$$

To the right of $\varphi = 0$ the space is free of charge so that the potential gradient is constant and equal to the value at $\varphi = 0$ obtained by taking the derivative of equation (39). This value is

$$\frac{d\varphi}{d\sigma} = -\frac{4\sqrt{2}\beta^{1/4}}{3}. \quad (41)$$

Concerning Completeness

We may now review our work and see that no possible space charge distributions can have been omitted. Starting from the fundamental equation (6) we obtain equation (7) with an undetermined integration constant. Setting this constant equal to zero we could integrate once more, obtaining a solution formally identical with Child's equation. If we supposed that the cathode plane defined by the Child's solution lay to the right of the initial plane, then the only freedom left in the solution was represented by Z, the fraction of current passing through the plane. All physically sensible values of Z, i.e., 0 to 1, are included in the solutions. If the cathode is assumed to lie to the left of the plane, then Z must equal 1 and the solution which arises is given by $\alpha = 0$ in equations (23) or (26), or $\beta = 0$ in equation (35)—that is, a D solution. For a negative value of the constant, a further integration gave only two possibilities. Each of these was investigated for all possible values of the constant. A similar statement is true for positive values of the constant.

As was stated in the text, space charge distributions corresponding to injection from both bounding planes can be handled in formally the same way as injection from one plane; therefore we may conclude that all solutions to the problem given by specifying the boundary conditions on two planes, subject to the assumptions represented by equation (6), have been determined.

Equations for Boundary Curves

For convenience we append a table of equations for the limiting curves occurring in the figures.

Symbol	Equation numbers	σ	γ	Z	$Z\gamma$
a	31	$(1+\varphi^{1/2})^{3/2}$	$(1+\varphi^{1/2})^3$	1	$(1+\varphi^{1/2})^3$
b	17, 27	$1+\varphi^{3/4}$	$(1+\varphi^{3/4})^2$	1	$(1+\varphi^{3/4})^2$
c	15, 16	$(1+\varphi^{1/2})^{3/2}2^{-1/2}$	$(1+\varphi^{1/2})^3/2$	$\frac{2\varphi^{1/2}}{(1+\varphi^{1/2})}$	$\varphi^{1/2}/(1+\varphi^{1/2})^2$
d	28	$(1+2\varphi^{1/2})\sqrt{1-\varphi^{1/2}}$	$(1+2\varphi^{1/2})^2(1-\varphi^{1/2})$	1	$(1+2\varphi^{1/2})^2(1-\varphi^{1/2})$
e	29	$(\varphi^{1/2}+2)\sqrt{\varphi^{1/2}-1}$	$(\varphi^{1/2}+2)^2(\varphi^{1/2}-1)$	1	$(\varphi^{1/2}+2)^2(\varphi^{1/2}-1)$
f	18	∞	∞	0	$\varphi^{3/2}$

The further relationship

$$\alpha = \varphi_{\min.} = \varphi(1+\varphi^{1/2})^{-2} \tag{30}$$

holds for the *a* curve.

A Carrier Telephone System for Toll Cables *

By C. W. GREEN and E. I. GREEN

A new 12-channel carrier telephone system for existing cables is described. This system, which incorporates a number of interesting departures from the previous carrier art, is now being manufactured in considerable quantities to meet increased traffic requirements.

AN important advance in the art of carrier telephony has been made by the development of a new 12-channel system, known as the type K, for toll telephone cables of existing type. It is applicable both to cables installed underground, and also to aerial cables, for which the wide range of temperature variation introduces quite difficult transmission problems. Field trials on cables previously installed between Toledo and South Bend have been successful, and the system is now being manufactured to meet field demands.

This new development is an outgrowth of the experiments at Morristown, New Jersey, described by Messrs. Clark and Kendall before the American Institute of Electrical Engineers in 1933,[1] and the essential principles of the new system were included in those experiments. The earlier work dealt, however, with cable specially designed for carrier operation, and only underground cable was experimented with. As that work drew to a close, it became clear that because of general economic conditions several years would elapse before the Bell System would require any substantial increase in toll facilities. Hence this early system was not put into commercial form, but work was continued to determine the extent to which carrier could be applied to existing cables, of which more than 15,000 miles were available for such use. Serious problems of cross-talk at high frequencies had to be reckoned with. A more serious problem, however, was that of maintaining stability of transmission, since with aerial cable, which comprises about two-thirds of the existing cable mileage, the total variation in attenuation, due to temperature variation, is about three times that for underground cable, and the rate of variation not infrequently is several hundred times as great.

In spite of these and other difficulties, the capabilities of the present system go far beyond those of previous systems. As a develop-

* Presented at Winter Convention of A.I.E.E., New York, N. Y., Jan. 24–28, 1938.

[1] For references see end of paper.

ment objective the maximum length was taken as 4000 miles, with as many as five separate systems linked together. On the basis of results thus far obtained it is expected that for these exacting conditions the performance with respect to crosstalk, noise, transmission stability, width of voice band and other characteristics will equal or exceed that of previous facilities for much shorter distances.

Superior performance has been achieved without material effect on the cost of the system. For distances of a few hundred miles, on moderately heavy traffic routes, it will provide telephone circuits at a much lower cost than previous facilities. The minimum distance for which the system will be useful may be less than 100 miles.

Interesting features of the new system are:

(1) A line of very high attenuation, requiring high-gain repeaters spaced at approximately 17-mile intervals. This would mean, for the maximum distance for which the system is designed, more than 200 repeaters in tandem.

(2) The use in the repeaters of the negative feedback principle of amplification to obtain the requisite stability and freedom from modulation.

(3) Small auxiliary repeater stations, established between existing voice-frequency repeater stations, housing equipment which can be left for considerable periods of time without attention.

(4) A system of transmission regulation whereby huge variations of attenuation, differing at each frequency, are automatically equalized to a high degree of accuracy.

(5) New methods of crosstalk and noise reduction. Small adjustable mutual inductance coils are connected between carrier pairs to balance out the crosstalk. The noise is kept at an extremely low level to permit the high gains.

(6) Channel terminal equipment designed so that it may be used in other types of carrier systems, thus simplifying development and manufacture, and facilitating the interconnection of different types of systems.

(7) Speech bands considerably wider than those of existing facilities. The increase is obtained by spacing the channels at uniform 4000-cycle intervals, and employing channel band filters containing quartz crystal elements.

(8) High-speed transmission, which is of considerable value from the standpoint of minimizing delays and echoes.

A general description of the system is presented herein, and the different parts are taken up in greater detail in other papers.

General Considerations

The type K system, the elements of which are illustrated schematically in Fig. 1, operates on a "four-wire" basis, using the same frequency range, but different electrical paths, for opposite directions of transmission. Thus it differs from open-wire carrier systems, for which the line is not suitable for four-wire operation, and which therefore require complicated and expensive filters to separate the different frequency bands used for transmission in opposite directions. A high degree of shielding between the two cable paths is necessary to avoid the effects of near-end crosstalk, which would be serious because of the large level differences existing at the repeaters and the terminals. On routes where two or more cables exist, such shielding is obtained by employing two separate cables, with transmission in one direction only, in each section of cable. On single-cable routes, a similar arrangement is obtained by adding a small cable. Where there is no cable, two small cables may be provided. Also, satisfactory shielding between the carrier pairs used for opposite directions of transmission has been obtained in short experimental lengths of cable by the use of a layer shield.

Frequency Allocation

In contrast to the original Morristown system, which gave nine one-way channels per pair in the range from 4 to 40 kilocycles, the type K system has twelve channels in the range from 12 to 60 kilocycles. As shown in Fig. 2, the frequency range of the type K system is roughly double that of preceding open-wire carrier systems.[2] The choice of 12 and 60 kilocycles as the lower and upper frequency limits was governed by economic considerations, and there is nothing technically insurmountable either in going to considerably higher frequencies or in utilizing the lower frequency range, which is now idle except for the use of the d-c. path for purposes of transmission regulation and fault location. Important factors influencing the selection of the upper frequency are the crosstalk, which depends on the number of pairs utilized for carrier in one cable and the extent to which special crosstalk balancing means are used, and the attenuation, which largely controls the spacing between repeaters. Factors affecting the lower limit include the difficulty of maintaining accurate transmission regulation over the whole frequency range, and the design of the repeater, which becomes harder as the ratio of maximum to minimum transmitted frequency is increased.

The frequency range between 12 and 60 kilocycles accommodates 12 speech channels, each occupying a gross band of 4 kilocycles. The

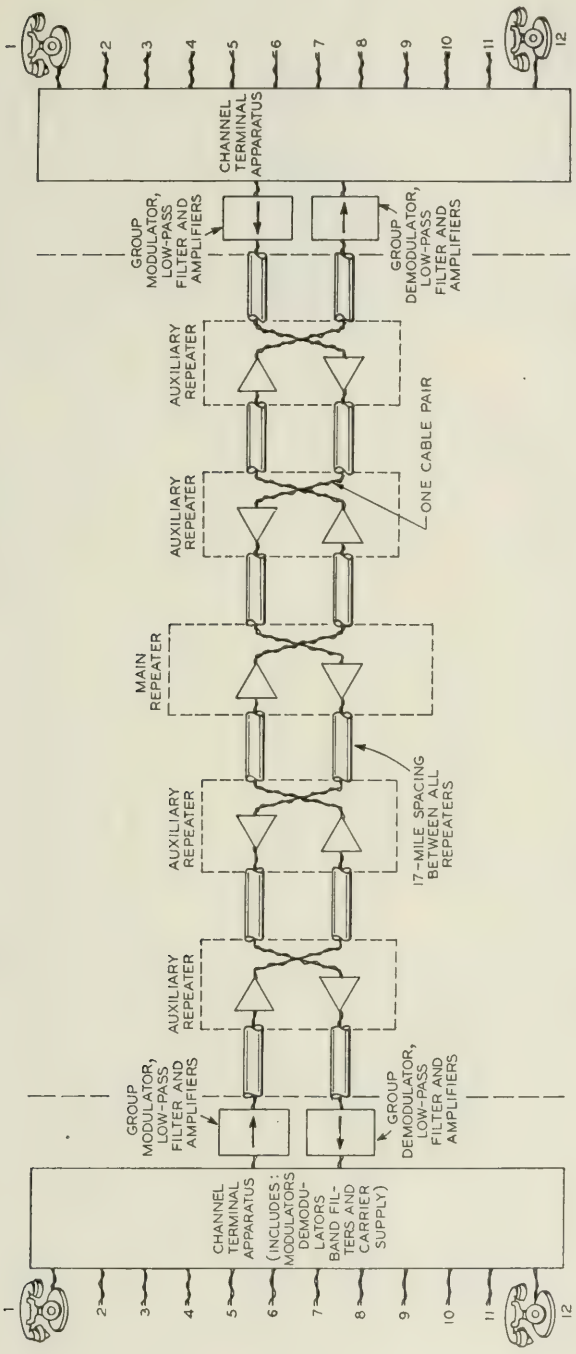


Fig. 1—Schematic of Type K system.

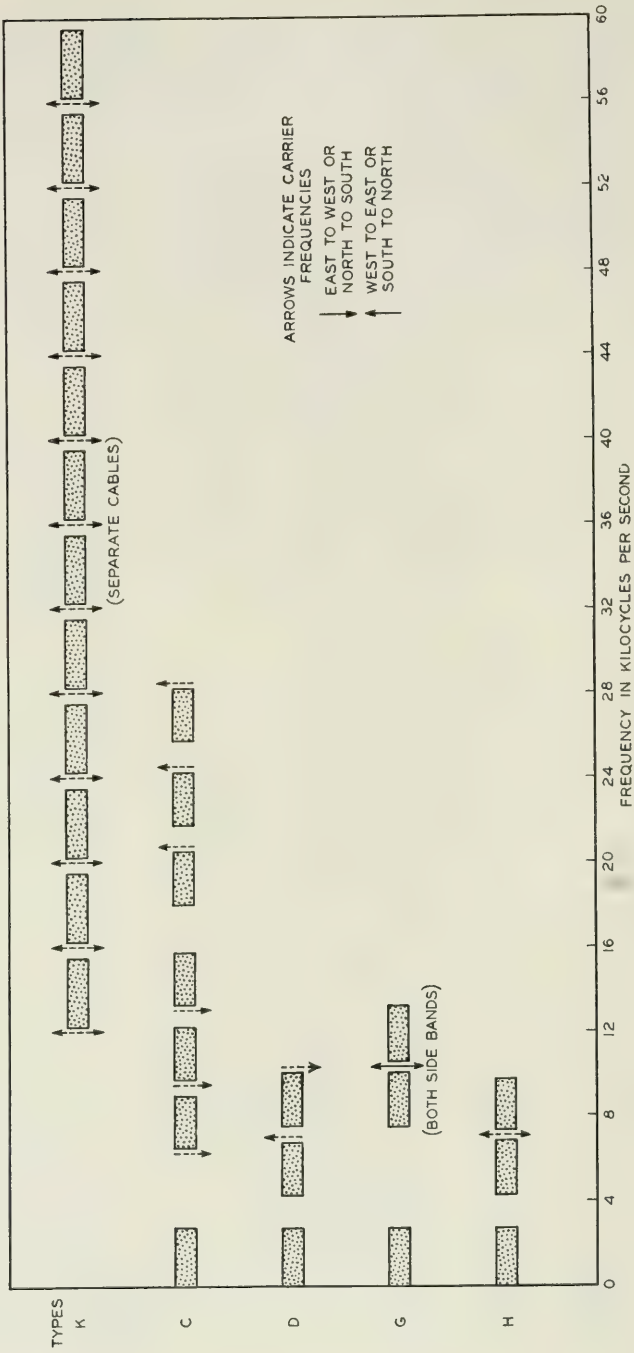


Fig. 2—Frequency allocation of carrier telephone systems.

single sideband method of transmission is employed, with carrier frequencies suppressed. The choice of a group comprising 12 channels was influenced not alone by the requirements of the type K system itself but also by those of other broad-band systems. From the earliest stages of the broad-band development it was recognized that there would be considerable advantage from the standpoints of flexibility of interconnection, of minimum development effort, and of large scale production of equipment units, if the designs of different broad-band systems could be so coordinated as to enable the same design of channel terminal equipment to be employed for each. A common 12-channel terminal unit developed for this purpose is used in the type K system.

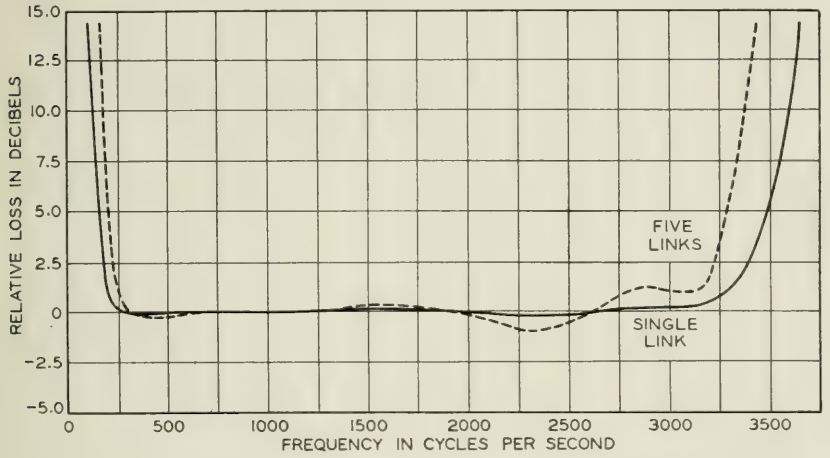


Fig. 3—Transmission frequency characteristics of overall circuit.

The spacing of the channels in broad-band systems is important from the standpoint of the channel selecting circuits and the width of the derived voice circuit. As discussed in a recent article, a uniform 4000-cycle interval has been adopted for the different channels of all broad-band systems.[3] The speech band width obtained with this spacing is in keeping with recent improvements in telephone instruments and other parts of the telephone plant. Overall transmission-frequency characteristics for a single link and a five-link connection are shown in Fig. 3.

Cable Attenuation

The type K system is designed to be applied to the No. 19 AWG (0.9 mm.) pairs commonly found in existing cables. (The Morristown

system used 16-gauge pairs.) Because the conductors are small and closely spaced, with paper and air dielectric, the attenuation of a non-loaded 19-gauge pair at the frequencies involved is inherently high, as will be seen from Fig. 4. Because of the high attenuation, the repeaters must be placed much closer together than is necessary for voice-frequency cable circuits. Fortunately this effect is partly offset by the fact that it is possible, as discussed later, to use higher gains in the carrier repeaters.

The cable pairs exhibit the rise in attenuation with frequency which is familiar in most transmission circuits. This effect is brought about largely by the increase in conductor resistance, due to skin effect, and

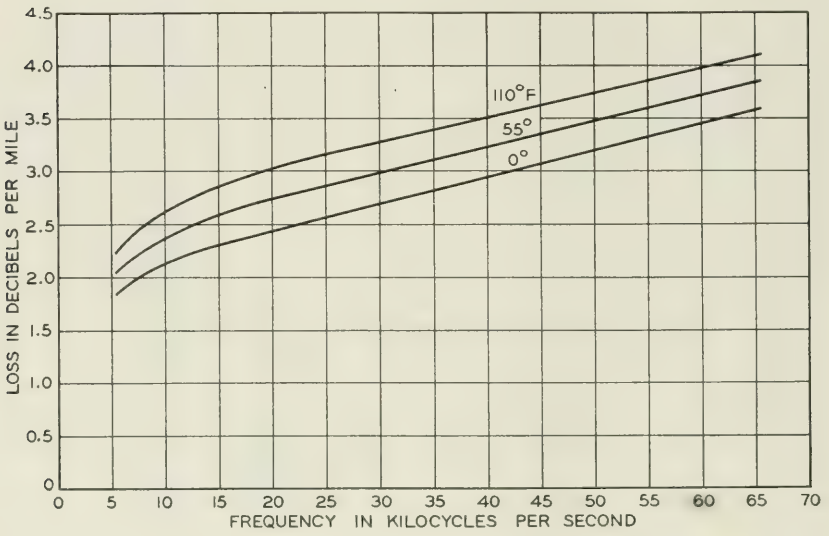


Fig. 4—Attenuation of 19-gauge non-loaded cable pair.

the increasing dielectric losses. More important than this, however, is the fact that the resistance of the wires and the other "constants" of the cable pair undergo variations with temperature, which in turn affect the attenuation. The magnitude of the result for a representative non-loaded 19-gauge cable pair is illustrated by the curves of Fig. 4, which show, respectively, the attenuation for an average temperature, assumed to be 55° F., and for 0° F. and 110° F. The latter values, often taken as the extremes of annual variation for an aerial cable, are in fact frequently exceeded. One reason for this is that when the sun is shining directly on an aerial cable, it may assume a temperature from 15° to 25° above that of the ambient air. The

range of temperature variation (and attenuation variation) for an aerial cable may be half as much in one day as in an entire year. For an underground cable, changes of temperature occur quite gradually and the total annual variation is about one-third of that for an aerial cable.

These relations between the attenuation of a cable pair and the frequency and temperature are of fundamental importance in the design of the type K system. First of all, since the attenuation at 60 kilocycles is about 4 db per mile, the total attenuation for a cable circuit of the length used in designing the type K system, i.e., 4000 miles, would be approximately 16,000 db. This must be offset by a corresponding gain.

In the next place, differences in the attenuation at the different frequencies would, if uncorrected, become so great that signals of the less attenuated channels would overload the repeaters, while those of the more attenuated channels would drop down into the noise region. Hence, each repeater must be given a gain-frequency slope which is complementary to the attenuation slope of the line.

Finally, the changes of transmission due to temperature variations and other causes must be compensated so precisely that the net variation in each channel is held within very narrow limits. The method of doing this is explained later. Here it is interesting merely to consider the magnitude of the problem. For the top channel, assuming a 4000-mile circuit, the annual variation in attenuation of an aerial cable pair might be approximately 2000 db. The systems thus far installed have, of course, been limited to much shorter distances than this.

Even if the change of attenuation with temperatures were related to frequency by a simple law, correct compensation over the frequency range would be far from easy. To a casual inspection the differential between any two curves of Fig. 4, for example those for 55° F., and 110° F., will not appear serious. This differential, which becomes very large for a long circuit, is a complicated function of the frequency.

The attenuation differential with temperature can be considered as made up of two components, one which is independent of frequency and another which varies with frequency. The former component, which is much the larger, requires a gain adjustment which is uniform or flat over the frequency range of the system. The latter component is frequently referred to as the "twist." For the range from 12 to 60 kilocycles, the maximum change of attenuation with temperature occurs near 28 kilocycles. Hence this frequency has been used as a datum point in determining the twist. The shape of the twist com-

ponent is apparent from Fig. 5, which shows the net loss per mile at temperatures of 0° F., and 110° F., assuming that the attenuation has been equalized so as to obtain a flat characteristic at 55° F., and that the gain is then adjusted so as to hold the transmission constant at 28 kilocycles as the temperature varies. Although the twist is small enough so that it need not be corrected at each repeater, it is too large to be allowed to accumulate over a very long distance.

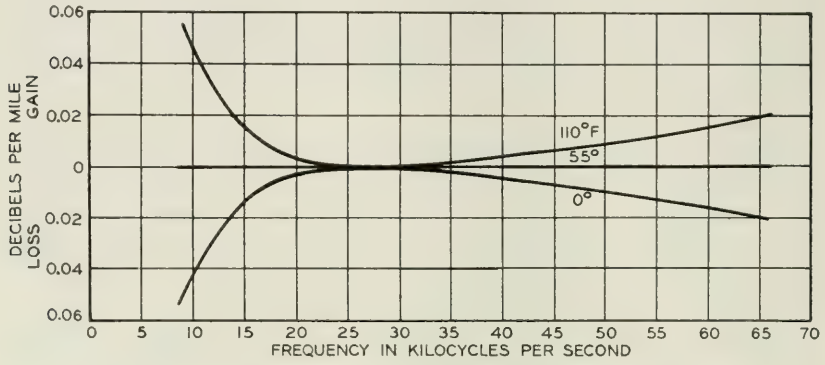


Fig. 5—Twist characteristics of 19-gauge non-loaded cable pair.

Crosstalk

As noted above, crosstalk between opposite directions of transmission is avoided by using two separate cables (or shielded compartments in the same cable). To prevent crosstalk in offices, special measures are employed. There remains the problem of "far-end" crosstalk between pairs in the same cable which transmit in the same direction. The pairs are packed closely together, and substantial crosstalk occurs between them because of small departures from symmetry and slight imperfections of twisting. However, by abandoning the use of phantoms and by connecting small adjustable mutual inductance coils between each carrier pair and every other carrier pair, sufficient crosstalk reduction is obtained to permit transmission up to 60 kilocycles on a substantial number of pairs.[4] The scheme is illustrated in Fig. 6.

In the original Morristown cable, the crosstalk was reduced in part by separating the 16-gauge carrier pairs from one another by 19-gauge quads which served as spacers. With existing cables, however, the use of spacers would be impracticable since this would require resplicing the cable at every joint, and therefore reliance must be placed largely on balancing. Since the number of combinations to be balanced increases approximately as the square of the number of pairs employed for carrier, the number of balancing coils required for even a moderate

complement of carrier pairs becomes quite large. With 40 carrier pairs, for example, the number of coils required is 780. The balancing coils are mounted on panels as shown in Fig. 7 and are connected together in a crisscross arrangement. Each repeater section is balanced separately, the balancing panels being located in the repeater station.

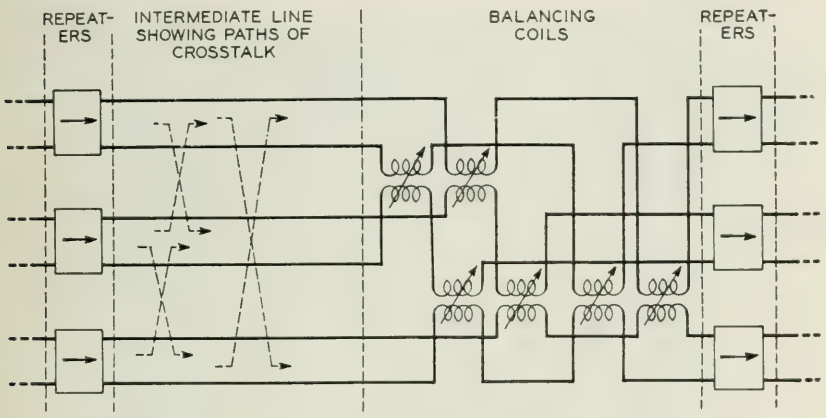


Fig. 6—Method of balancing out crosstalk.

Other measures are necessary to supplement this balancing technique. To reduce the crosstalk coupling, and also to average the transmission characteristics of different pairs, the carrier pairs in different quads of an existing cable are respliced about every mile so as to approach random splicing. The crosstalk coupling between the two sides of a quad is reduced by test splicing at the middle of a repeater section and the quads are split at repeater points. In a new cable, of course, the desired splicing arrangements are introduced at the time of installation. The carrier pairs are also transposed from one cable to the other as indicated in Fig. 1. This avoids interaction crosstalk that would take place, through the medium of the voice-frequency pairs, between the high-level carrier outputs on one side of a repeater station and the low-level inputs on the other side. There is of course, a similar effect between carrier pairs at any point in a repeater section. This is much less serious since no level difference is involved, but it does tend to limit the effectiveness of the balancing over a range of frequencies.

Reflections resulting from impedance irregularities reverse the direction of propagation and therefore produce far-end crosstalk from near-end crosstalk. This crosstalk cannot readily be balanced out

over a range of frequencies. To avoid it, it is necessary that the impedances of successive lengths of cable pair be substantially uniform and also that the impedance of the equipment be closely matched to the characteristic impedance of the cable pair.

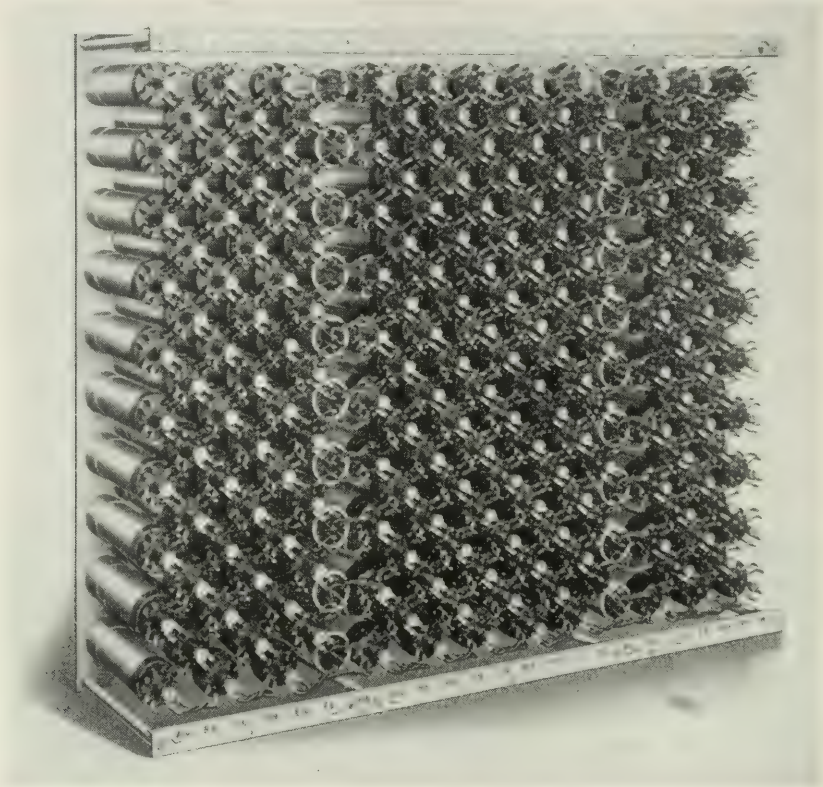


Fig. 7—Crosstalk balancing panel.

Noise

The cable pairs are fairly well protected by the lead sheath from external electrical disturbances. However, high-frequency noise originating in the voice-frequency repeater stations due to relay operation, etc., would, unless prevented, enter the cable over the voice-frequency pairs and thence would be induced in the carrier pairs. Such noise would be excessive at the low-level carrier inputs. It is avoided by connecting, in each voice-frequency pair in the "low level" cable on each side of a voice-frequency repeater station, a coil which suppresses longitudinal noise currents. Similarly, it is necessary to

keep high-frequency noise from entering the cable where open-wire pairs tap into it and frequently also where branch cables are connected. For this purpose simple noise suppression filters are employed. With these and other measures the noise on the carrier pairs can, at the highest frequencies where the amplification is greatest, be brought within a few db of the basic noise due to thermal agitation of electricity in the conductors themselves.

Velocity of Transmission

Voice waves travel through loaded cable circuits at from 10,000 to 20,000 miles per second, the higher speed being used for the longer circuits. On very long connections, even if echo suppressors are employed this velocity results in transmission delays which introduce difficulties in conversation.[5] The use of non-loaded conductors for the type K systems results in an overall velocity of transmission of about 100,000 miles per second, a speed so high that such difficulties are greatly reduced and satisfactory telephone conversations are possible over the longest distances for which connections may be required.

Repeaters

Since the noise level in the cable circuits can be made quite low, the carrier currents may be permitted to drop to levels below those used on voice-frequency circuits or on open-wire carrier circuits, and the repeaters may have higher gains. In the cable carrier system, the noise has been so reduced that the level of the top channel at the repeater input may on the average be dropped about 60 db below the voice level at the transmitting switchboard. The amplifier gains at the top frequency range from about 50 to 75 db, and the output level of each of the 12 channels is about 10 db above that at the switchboard. The average repeater spacing is about 17 miles.

The tube which was developed for the gain stage of the amplifier is a pentode with indirect heater. The heater requires a potential of 10 volts and a current of 0.32 ampere and the plate 150 volts. The tube in the power stage is similar in type but requires a heater current of 0.64 ampere at 10 volts. With this power tube a feedback of about 40 db has been found to provide a satisfactory reduction of interchannel modulation.[6] Both tubes were designed to have long life with very reliable performance.

Description of Amplifier

Each repeater comprises two amplifiers of the type illustrated in Fig. 8. A schematic diagram of the amplifier circuit is shown in

Fig. 9. Three stages with impedance coupling are used and the feedback circuit is connected between the plate circuit of the last tube and the grid circuit of the first. The amount of gain and the slope of the gain-frequency characteristic are controlled by the condensers and line equalizer in the feedback circuit.

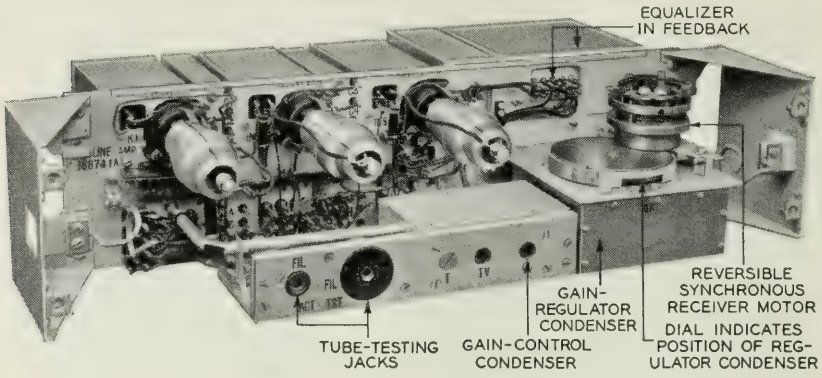


Fig. 8—Line amplifier.

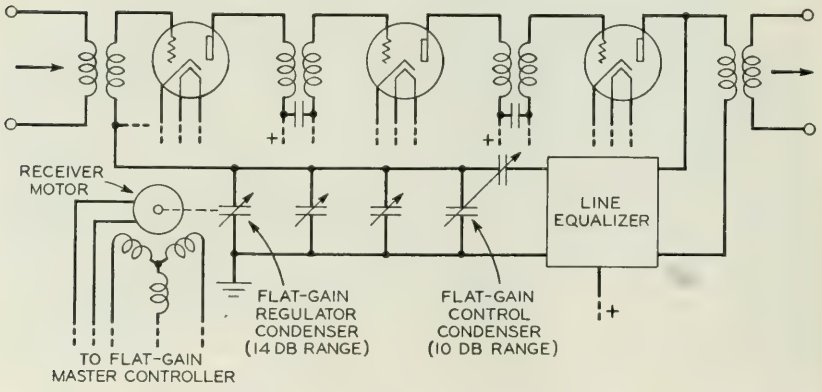


Fig. 9—Schematic of line amplifier.

The line equalizer has the same attenuation slope as the line. In the introduction of this equalizer in the feedback circuit careful attention to phase shift requirements was required. Four types of equalizers are available, for different repeater spacings, to compensate for the cable distortion which occurs at a temperature of 55° F., additional means being provided to compensate for variations which occur as the cable temperature swings away from this value. The solid curve of Fig. 10 shows a repeater gain characteristic with one of these equalizers in the feedback circuit.

The correction introduced by the line equalizers is subject to errors which, although small at each repeater point, become important for a moderate length of system. Supplementary equalizers have been designed to correct for these. Two types of deviation are considered,

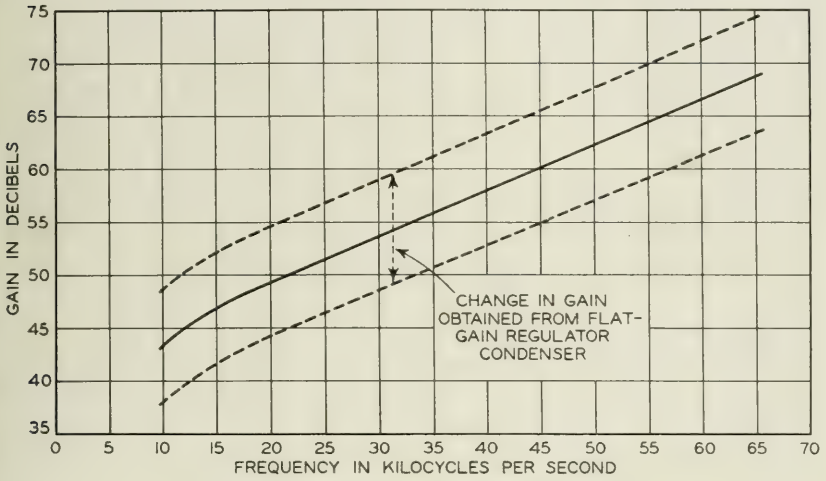


Fig. 10—Amplifier gain characteristics.

that of the cable and that of the amplifier. As the characteristics of cables manufactured at different times show slight departures from one another, two shapes are required to correct their deviations. There is one correction for concave deviations and another for convex.

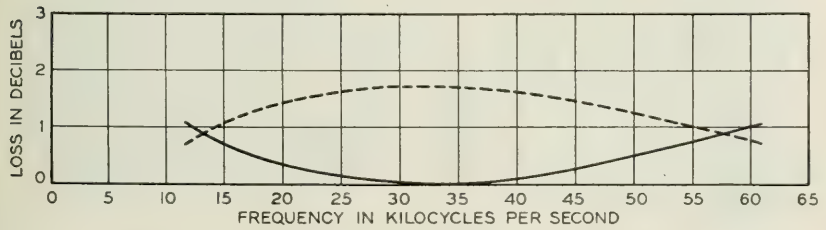


Fig. 11—Characteristics of cable deviation equalizers.

The amplifier requires but one type of correction. The characteristics of these equalizers are shown in Figs. 11 and 12. The equalizers for amplifier deviations are used about every 10th repeater and those for the cable deviations, at distances of 300 to 400 miles. At normal temperature (55° F.), the correction applied by these networks will,

for a 500-mile system, result in a frequency characteristic which is flat to within less than 2 db over the range from 12 to 60 kilocycles. For a longer circuit further corrective measures will be provided.

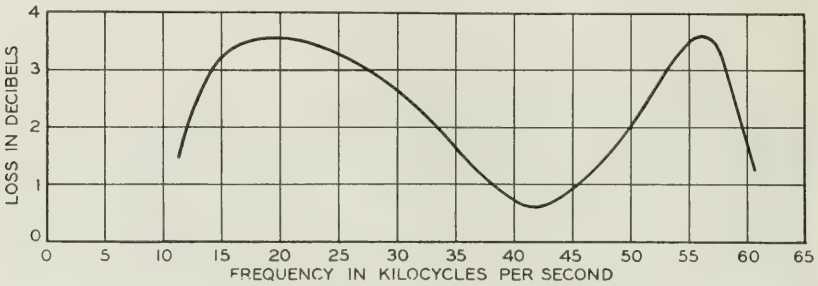


Fig. 12—Characteristics of amplifier deviation equalizer.

Regulation for Temperature Effects.

The method adopted for controlling the repeater gain to compensate for temperature changes is similar to that which has been found satisfactory for voice-frequency cable circuits. This is the pilot wire method in which a pair of cable conductors extending over the section

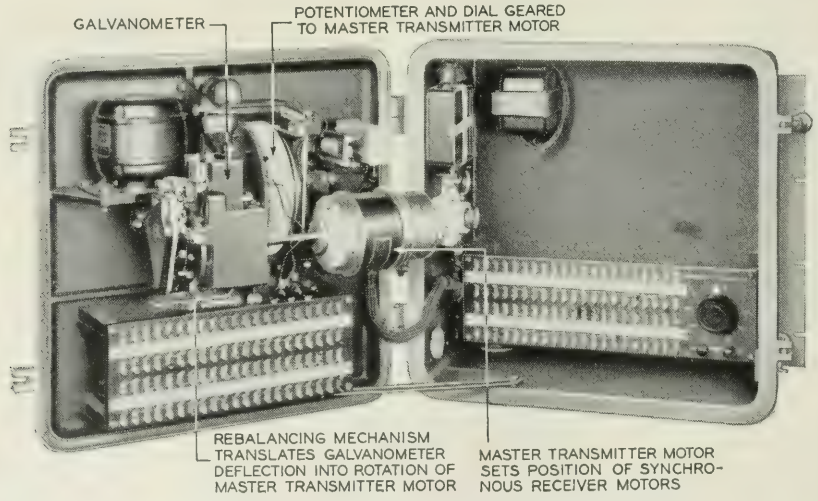

GALVANOMETER
POTENTIOMETER AND DIAL GEARED TO MASTER TRANSMITTER MOTOR
REBALANCING MECHANISM TRANSLATES GALVANOMETER DEFLECTION INTO ROTATION OF MASTER TRANSMITTER MOTOR
MASTER TRANSMITTER MOTOR SETS POSITION OF SYNCHRONOUS RECEIVER MOTORS

Fig. 13—Flat gain master controller.

to be regulated forms one arm of a Wheatstone bridge. This bridge is designed for automatic self-balancing and the mechanical motion required for establishing the balance has been made to adjust the gain of the amplifiers. The d-c. resistance of the pilot wire gives an

accurate indication of the temperature of the carrier pairs which determines their attenuation to a close approximation. The motion of the bridge mechanism is communicated to the repeater amplifiers by means of self-synchronizing motors, a master motor being associated with the bridge and an individual motor with each amplifier.

With aerial cable a flat gain correction must be made at every repeater. With underground cable the flat gain correction may be

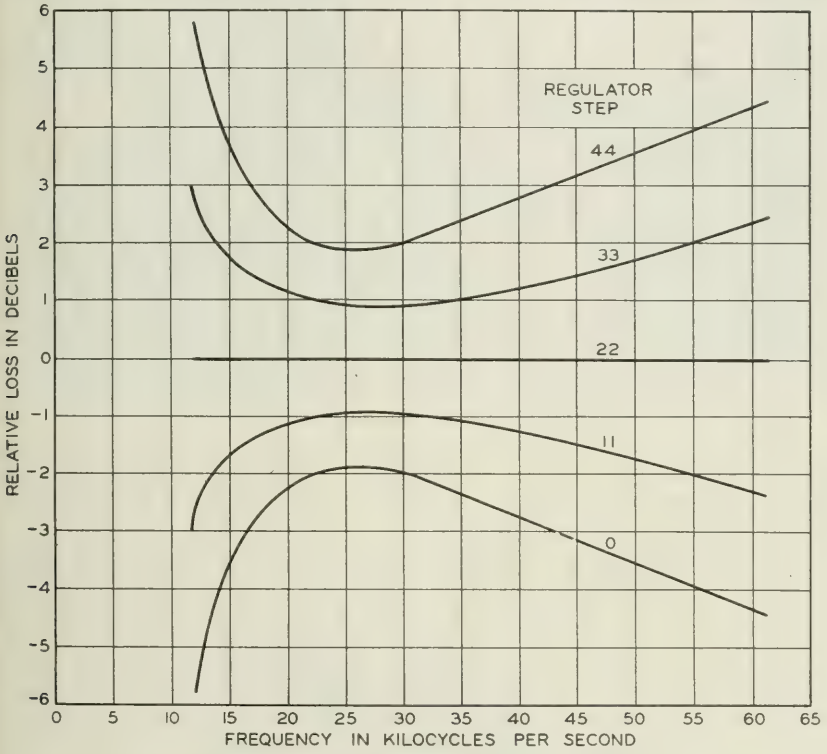


Fig. 14—Characteristics of twist regulator networks.

omitted at some repeater points. Figure 13 shows a master controller with its galvanometer, driving motor, and self-synchronizing motor. The air condenser in the feedback circuit of Fig. 8 makes this correction. The small self-synchronizing motor which may be seen in this figure is geared to the condenser and it moves the air condenser corresponding to the motion of the master motor. The resulting change in repeater gain is virtually the same for all frequencies in the transmitted band. In Fig. 10 the repeater gain is plotted against frequency for three angular positions of the condenser.

As was mentioned earlier, an additional correction for the residual effect or "twist" is required about every six repeaters to supplement the flat gain adjustment. This distortion is a function of frequency and has been found to vary from cable to cable. A network the characteristics of which are shown in Fig. 14 has been developed to meet this condition. Certain fixed resistances in the network are selected to correspond to the length and twist characteristic of the cable section considered. A variable resistance in the network is adjusted automatically using a control similar to the flat gain regulator. Figure 15 gives the transmission characteristics of a 150-mile regulator-controlled circuit under two temperature conditions.

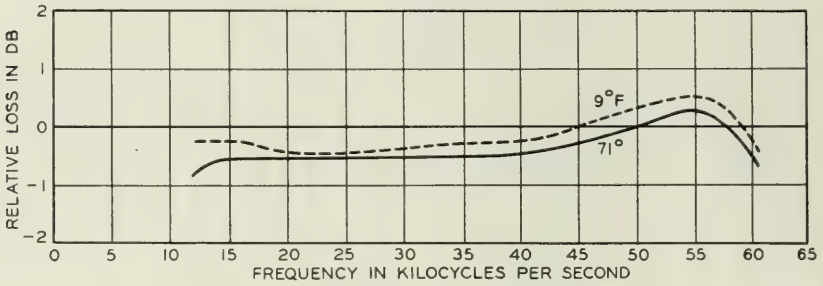


Fig. 15—Overall transmission-frequency characteristic of 150-mile line.

Auxiliary Repeater Stations

Cable carrier systems are expected to be used largely on existing toll cable routes which now carry voice-frequency circuits. The average spacing of the stations housing the voice-frequency repeaters on these routes is about 50 miles. The same buildings with their power plants will also care for the cable carrier repeaters. Since the maximum spacing for the carrier repeaters is about 19 miles, additional carrier repeaters must be provided at intermediate stations (two is the usual number). The various design features of the equipment to be located in these stations have been made the subject of extensive development work and field tests. These stations are designed to function with a minimum of attention and are visited at intervals for routine testing work or as required by some emergency, but resident maintenance forces are not planned for them. The present equipment is expected to be suitable not only for auxiliary stations on existing cable routes but also for cases where a greater spacing than 50 miles between the attended stations may be desired on new routes.

A voice-frequency repeater station for a single cable and a cable carrier auxiliary station are shown to approximately the same scale

in Figs. 16a and 16b, respectively. Many of the existing voice-frequency stations are even larger than that shown in Fig. 16a. The auxiliary building shown in Fig. 16b has about 600 feet of floor space

Fig. 16*a*—Voice frequency repeater station on single cable route.

with a ceiling height sufficient to take care of 11′6″ relay racks. This building will house 100 repeaters with necessary auxiliary equipment, thus providing ultimately for a total of 1200 carrier circuits. The interior of a typical auxiliary station is shown in Fig. 17.

Fig. 16*b*—Auxiliary cable carrier repeater station.

The main power plant for the repeaters consists of a 152-volt storage battery, which is continuously floated across a grid-controlled rectifier fed from the 60-cycle power mains. The voltage of the entire

battery supplies the plate voltage for the tubes. Each amplifier requires about 22 volts for the tube heaters and this is obtained by dividing the battery into seven sections, each section supplying several amplifiers in parallel. Additional power supplies of 55-volt alternating

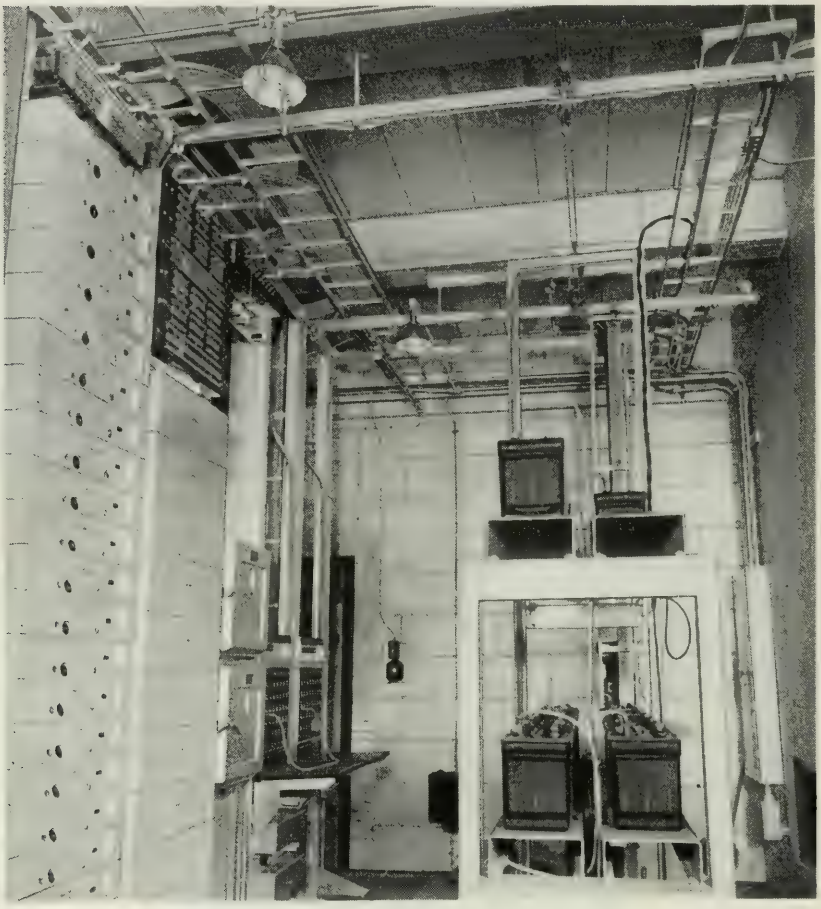


Fig. 17—Interior of auxiliary repeater station.

current and 140-volt direct current are required for the regulator system.

In the station, there are alarm circuits, which signal the nearest attended office if trouble develops. There are alarms for blown fuses, high or low battery voltages, power failures, etc. A telephone order wire to the nearest attended station is provided for the maintenance force.

In addition to the line amplifiers with their regulating equipment, there are racks mounting the crosstalk balancing coils. There are also sealed terminal units between the outside cable or the balancing units and the office cable. These furnish access to the line or equipment through jacks.

Terminals

The minimum distance over which a cable carrier system can be operated economically is determined in large measure by the cost of the terminal apparatus. Hence, the field of usefulness of the system is greatly increased by keeping the terminal cost as low as is consistent with satisfactory performance. Numerous developments during the past few years in connection with modulation, filtering and methods of carrier supply have all contributed materially toward this end. At the same time, the standards of performance have not only been maintained, but in many respects substantially improved.

Channel and Group Modulation

In the design of the terminals for the type K system, a number of circuit arrangements were considered, the final choice being influenced to a considerable extent by the conditions imposed upon the filters. As noted above, the desirability of using the channel terminal equipment in other broad-band systems, such as those for open-wire or coaxial cable, was also an important factor. The circuit arrangements selected have a first stage of modulation which raises the voice frequencies of the 12 channels up to a range of 60 to 108 kilocycles. This range is favorable to the use of crystal type band filters,[7] which have transmission characteristics superior to the coil and condenser type and seem to be no more costly. For the type K system, a single stage of group modulation shifts the frequencies to the range required on the line, 12 to 60 kilocycles, and a similar stage at the receiving end returns them to the 60 to 108-kc. range. Other carrier systems will also use the 60 to 108-kc. channels and by group modulation shift them to the desired position in the frequency spectrum.

The band filter occupies a space on the relay rack equal to 1/8 of that required by the coil and condenser type which was used in the earlier model of this system. Its attenuation characteristic in the transmitting region is flat to within 1 db over a range of about 3100 cycles. Immediately outside of this range the attenuation rises very rapidly, thus permitting very efficient use of the frequency spectrum.

Another new device on the terminal is the copper-oxide unit used in the modulating process. These units are expected to show a stability

of the same order as that of coils and condensers, and require practically no maintenance as compared to vacuum tubes.

The translation of the channels from the 60 to 108-kc. range to the position required for cable carrier, 12 to 60 kilocycles, is made by a stage of group modulation. A copper-oxide group modulator is used and a carrier frequency of 120 kilocycles. The reverse of this process in a similar group demodulator at the receiving end steps the frequency back to its original range, 60 to 108 kilocycles. These processes of modulation take place at points of low-energy level in the circuit with a comparatively high level of carrier, so that the inter-channel crosstalk which results from unwanted products of modulation is unobjectionable. Low-pass filters are inserted after the group modulator and demodulator, and amplifiers with flat gain characteristics are supplied to raise the levels of the output currents of the group modulators or demodulators.

Carrier Supply

The carrier frequencies which are required at a terminal are obtained from the harmonics of a base frequency. The carrier supply system is common to as many as 10 systems in one office. This simplification was made possible by the selection of the channel frequencies as multiples of a base frequency, 4 kilocycles being chosen for this system. This base frequency is produced by an oscillator in which the control element is a tuning fork, the whole unit being designed to have the necessary output and frequency stabilities. The output of the oscillator is amplified and fed to a circuit which produces the desired harmonics. All of the carrier frequencies which are required for the different channels as well as for group modulation and demodulation are obtained from these harmonics. A small coil with a permalloy core is the important agent in this process.[8]

Failure of the 4-kc. supply, or failure of the 120-kc. supply used for group modulation, would cause failure in the channels of all systems operating from this supply. Provision is made for such a contingency by an emergency carrier supply which is automatically switched into service when the regular supply fails. This reserve source duplicates all of the parts of the regular supply, 4-kc. fork, amplifier, harmonic producer, and amplifier for the 120-kc. carrier.

Assembly

The different panel units which make up the terminal of a type K system are assembled on a functional basis with similar panels of other K systems, the channel modulator-demodulator panels in one

bay, the carrier supply in a second, the group modulator and demodulator in a third, etc. The compactness of the equipment makes it possible to mount the modulators and demodulators for 18 channels on one 11 ft. 6 in. bay 19 inches wide.

Signaling

The same type of ringdown signaling equipment is used with the channels of this system as with the voice-frequency toll circuits. A 1000-cycle tone, interrupted 20 times per second, is impressed on a channel terminal, modulated, and transmitted over the carrier system in the allotted channel band. At the far end, it is demodulated to operate the receiving end of the standard voice-frequency signaling circuit, or to be transmitted along an extended voice-frequency circuit to its terminal.

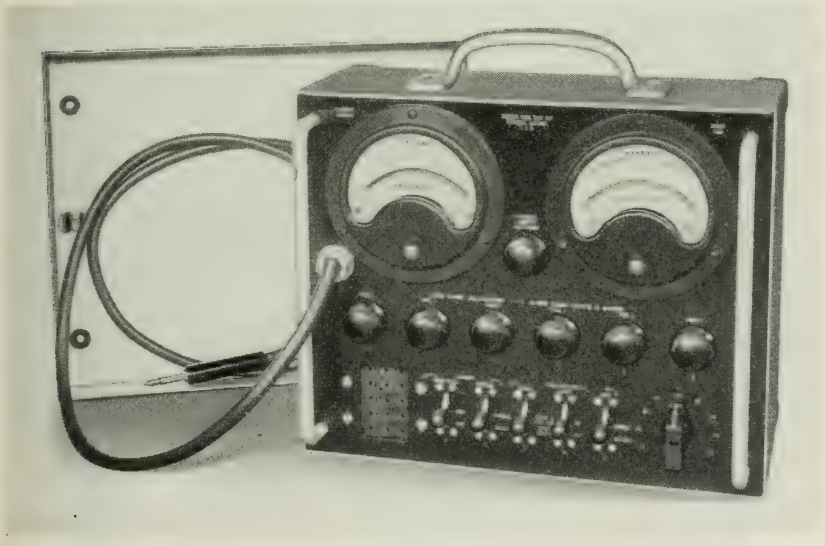


Fig. 18—Vacuum tube test set.

Telegraph and Program Applications

Voice-frequency telegraph can be superimposed on any of the carrier channels as is now done on the three-channel open-wire systems. Equipment is being developed to include a program channel on the cable carrier system. This will be done by devoting to the program circuit the frequency space occupied by two of the 4-kc. speech bands.

System Maintenance

Arrangements are provided whereby the tubes may be tested on a routine basis as has been done in voice-frequency practice. The amplifier panels, however, are provided with test jacks which are con-

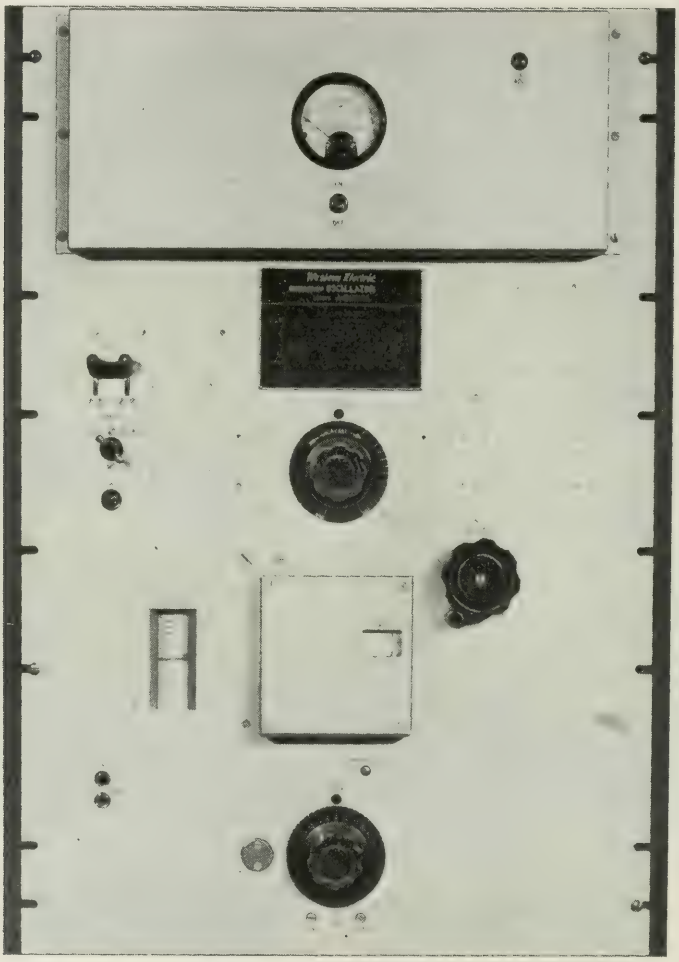


Fig. 19—Testing oscillator.

nected to resistances in the plate circuits. A reading of the voltage across the resistance gives a measure of the plate current for the associated tube without disturbing the amplifier performance while the amplifier is in service. A portable tube testing set, Fig. 18, has been designed for this measurement.

Provision is being made for the removal of an amplifier from an active circuit without interruption of service. A spare amplifier at each repeater station can be substituted for the active one by connecting it to jacks at the sealed terminal and operating associated relays to make a quick transfer.

Apparatus is also furnished which permits the substitution of a new link between attended points for one which develops trouble. A complete high-frequency circuit for each direction of transmission will generally be reserved as a spare. It can be substituted for any working

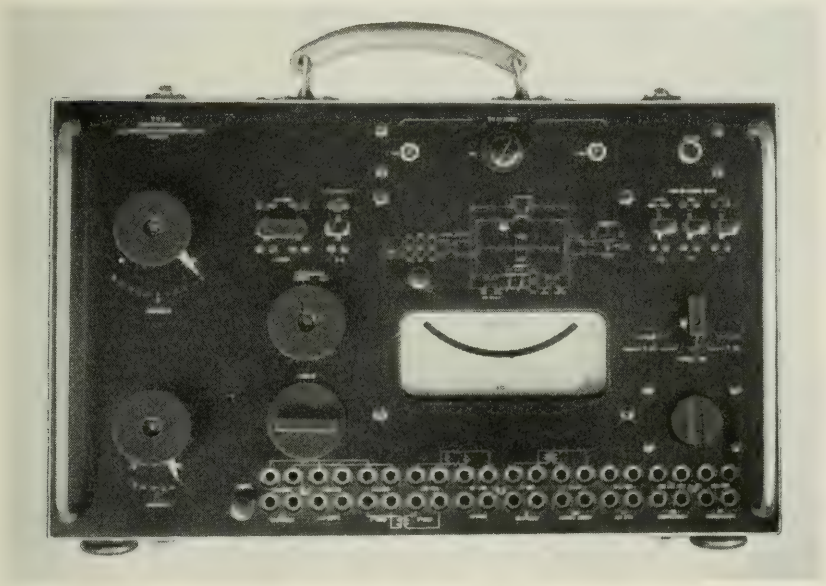


Fig. 20—Transmission measuring set.

high-frequency circuit without interfering with service by paralleling the transmitting ends of the spare and working circuits and patching the receiving ends through relays. The operation of a key controlling these relays substitutes the spare circuit for the working one with a transient disturbance of but 1 or 2 milliseconds.

Three pilot frequencies, 15.9, 27.9 and 55.9 kilocycles, which are produced at the transmitting terminals, may be used to check the levels at the main repeater points and the receiving terminals. This is done by means of a special testing circuit which can be bridged across a pair to detect the level of the pilots without interference to service.

A heterodyne oscillator having a frequency range from 60 cycles to 150 kilocycles has been developed for use in testing this and other carrier systems. Its frequency is calibrated at 60 cycles against the power mains and at 100 kilocycles against a quartz crystal. This oscillator is shown in Fig. 19. A portable test set, developed for measuring transmission gains and losses with high precision, is shown with the cover removed in Fig. 20.

Conclusion

The type K system makes possible the application of carrier to toll cables of existing type, whether installed underground or aerially. The blocks of 12 circuits each, which it furnishes, seem to be a convenient size for routes where large numbers of circuits are concentrated. It is to be expected, of course, that substantial modifications and improvements will be made in this system through further development effort. In its present form, however, it constitutes an important stage in the history of carrier development. Plans already under way call for the application of large numbers of such systems to meet rapid growth in long distance traffic.

This new system forms merely one phase of a concerted development effort on broad-band carrier transmission systems.[9, 10] There is every indication that, taken collectively, these broad-band systems will have far reaching effects upon the toll telephone plant of the Bell System. A transition is already under way from the time when carrier was used only on open wire, and comprised only a small part of the toll plant, to a time when carrier systems will furnish a major part of the toll circuit mileage of the Bell System. The type K system is clearly destined to play an outstanding part in this evolution of the toll plant along carrier lines.

References

1. "Carrier in Cable" by A. B. Clark and B. W. Kendall, *Electrical Engineering*, Vol. 52, page 477, July 1933; also *Bell Sys. Tech. Jour.*, Vol. XII, p. 251, July 1933.
2. "Carrier Systems on Long Distance Telephone Lines" by H. A. Affel, C. S. Demarest and C. W. Green, *Bell Sys. Tech. Jour.*, Vol. VII, pages 564–629, July 1928; also *Electrical Engineering*, Vol. 47, pp. 1360–1367, October 1928.
3. "Transmitted Frequency Range for Circuits in Broad Band Systems" by H. A. Affel, *Bell Sys. Tech. Jour.*, Vol. XVI, p. 487, October 1937.
4. A. G. Chapman, U. S. Pat. No. 1863651; M. A. Weaver and O. H. Coolidge, U. S. Pat. No. 2008061; M. A. Weaver, U. S. Pat. No. 2080217.
5. "The Time Factor in Telephone Transmission" by O. B. Blackwell, *A.I.E.E. Transactions*, Vol. 51, pages 141–147, March, 1932; also *Bell Sys. Tech. Jour.*, Vol. XI, pp. 53–66, January 1932.
6. "Stabilized Feed-Back Amplifiers" by H. S. Black, *Electrical Engineering*, Vol. 53, p. 114, January 1934.

7. "Electrical Wave Filters Employing Quartz Crystals as Elements" by W. P. Mason, *Bell Sys. Tech. Jour.*, Vol. XIII, p. 405, July 1934.
8. "Magnetic Generation of a Group of Harmonics" by E. Peterson, J. M. Manley and L. R. Wrathall, *Electrical Engineering*, Vol. 56, No. 8, p. 995, August 1937; also *Bell Sys. Tech. Jour.*, October 1937.
9. "Wide-Band Transmission in Sheathed Conductors" by O. B. Blackwell, *Bell Telephone Quarterly*, Vol. XIV, p. 145, July 1935.
10. "Systems for Wide-Band Transmission Over Coaxial Lines" by L. Espenschied and M. E. Strieby, *Electrical Engineering*, Vol. 53, pp. 1371–1380, October 1934; also *Bell Sys. Tech. Jour.*, Vol. XIII, pp. 654–679, October 1934.
11. "Modern Systems of Multi-Channel Telephony on Cables" by A. S. Angwin and R. A. Mack, *Journal of the Inst. of Electrical Engineers*, Vol. 81, No. 941, p. 573, November 1937.

Cable Carrier Telephone Terminals *

By R. W. CHESNUT, L. M. ILGENFRITZ and A. KENNER

This paper describes the circuits, performance and equipment features of the terminals of a new 12-channel carrier system for application to existing toll cables. The 12-channel group of terminal apparatus has been designed also to form a basic part of the terminals of other carrier systems now under development, such as the type J system for open wire and the coaxial system.

Introduction

ABOUT twenty years ago the first commercial carrier telephone system was installed between Baltimore and Pittsburgh. Until recently, telephone circuits were obtained by carrier methods largely on open-wire lines. The notable exceptions were on short deep sea submarine cables.[1, 2] Ten years ago, experiments were initiated which have now resulted in the design of a carrier system which can be applied with substantial economy to existing long distance toll cables on land. Its general features are described in another paper.[3] The present paper describes in detail the circuits and performance of the carrier terminals of this system.

General Features

The carrier system for existing cables, designated type "K," is designed to provide twelve telephone channels in the frequency range between 12 and 60 kilocycles, using one non-loaded 19-gauge paper insulated cable pair in each direction. Previous carrier systems employed for open-wire lines used vacuum tubes for the modulating or translating circuits and electrical filters composed of coil and condenser networks for separating the frequency bands associated with the respective channels. The terminals of the new type "K" system are simpler and yet provide improved performance by using copper oxide bridges for the modulation function and quartz crystal filters [4] for the separation of the individual channel bands.

The quartz crystal filter is economical only in a comparatively high-frequency range, necessitating the use of high intermediate frequencies. The high intermediate frequencies are reduced by a second stage of modulation to the desired range of frequencies for transmission over the

* Presented at Winter Convention of A. I. E. E., Jan. 24–28, 1938.

line. Copper oxide bridge circuits again are used for this group modulation stage. In all cases they are connected to suppress the carrier. To provide the various carriers required for modulation and demodulation, a carrier supply system has been designed somewhat along the lines of an office power distribution system using bus bars and protective arrangements for the various carriers. Each carrier supply system is capable of supplying as many as ten carrier terminals, or a total of one hundred and twenty two-way channels.

Because of the large number of circuits involved, every effort has been made to provide reliable operation of the carrier supply and common terminal equipment. The terminal and carrier supply equipment is designed to permit maintenance tests for checking the performance of amplifier tubes and to permit switching between regular and spare equipment without interruption of the large number of circuits involved.

The emphasis placed upon ease of maintenance and the necessity for more careful handling of higher-frequency circuits have resulted in new equipment design features. These include new cable terminals, new shielded office cabling, and panels arranged for front wiring and maintenance which are mounted on racks having wiring ducts at both edges of the bays. In the following sections a more detailed description is given of the circuits, their performance, equipment and maintenance features.

Circuits

The frequency allocation for one direction of transmission and a block schematic of one terminal are shown in Figs. 1 and 2, which supplement each other and need little explanation. The twelve voice bands shown at the left in Fig. 1 are modulated individually in the channel modems.* This forms a 12-channel block lying between 60 and 108 kc. which is then modulated in the group modulator by a 120 kc. carrier to move the block down in the range from 12 to 60 kc. for transmission to the distant terminal. On the receiving side the processes are reversed. One of the channels, as well as the group modem of Fig. 2, is presented in more circuit detail in Fig. 3. This shows the circuit from the point where the voice comes into the carrier system to the point where the twelve carrier sidebands go out onto the cable and vice versa.

At the left the four-wire terminating circuit serves, not only as a device to transform from a two-wire to a four-wire circuit, but also as a

* The term "modem" has been coined to mean a panel or equipment unit in which there is both a modulator and a demodulator to take care of both the outgoing and the incoming signal.

high-pass filter to eliminate, from the input to the carrier system, noises below about 200 cycles, such as telegraph harmonics, 20-cycle ringing, 60-cycle power, etc., which may be present on connected voice-frequency circuits. Otherwise these noise frequencies, which are below the voice range, would modulate and pass through the terminal to load unnecessarily the carrier repeaters along the line, as well as to interfere with the level indications of the pilot channels.

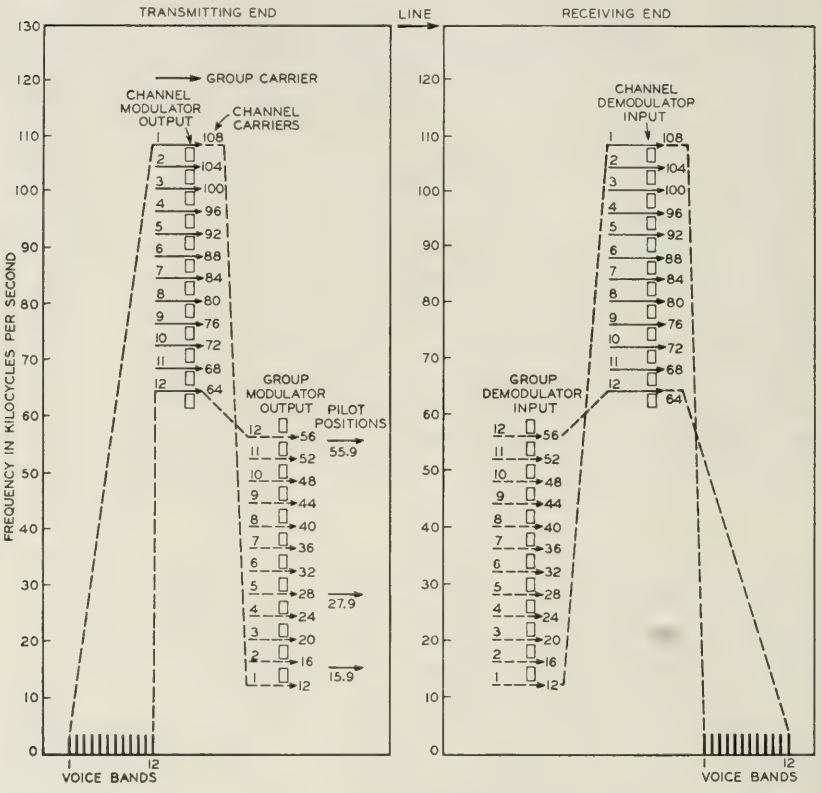


Fig. 1—Frequency allocation.

From the terminating equipment the circuit loops through jacks which have paralleled contacts for reliability. The level at this point is — 13 db compared with the transmitting toll switchboard, which level is expected to be generally used in the Bell System for all multi-channel carrier telephone systems. Then comes the channel modulator which consists of four copper-oxide discs, each three-sixteenths of an inch in diameter, potted in a small can. This makes a very simple and inexpensive modulator which is much more satisfactory than tubes.

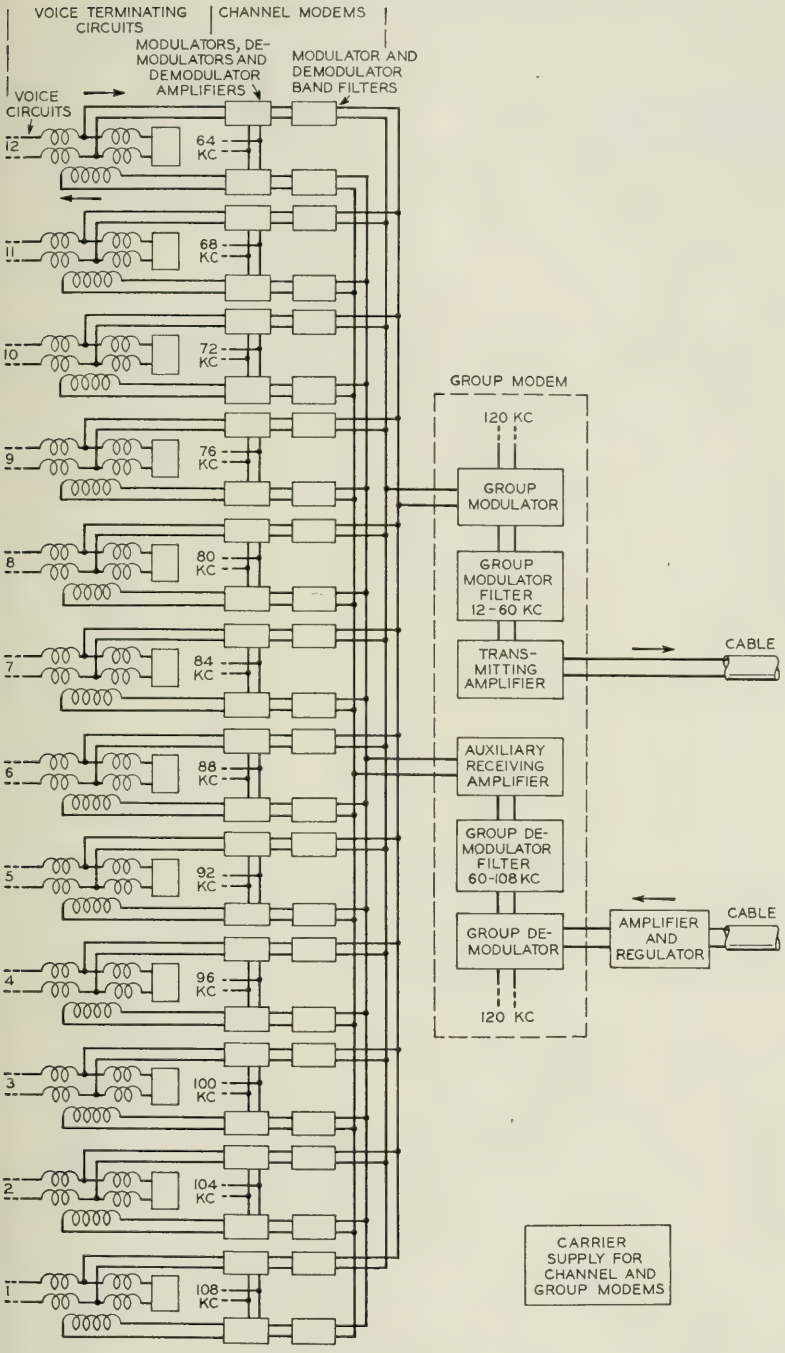


Fig. 2—Block schematic.

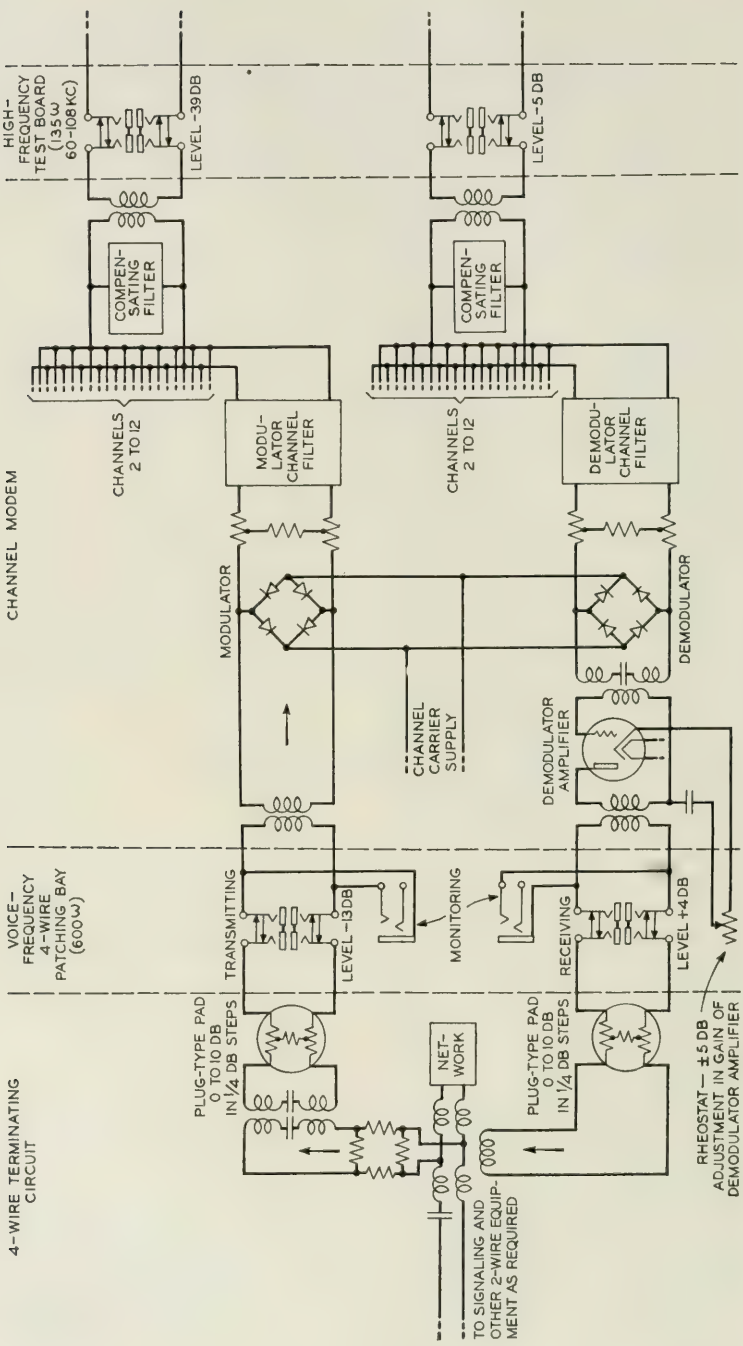


Fig. 3—Circuit schematic.

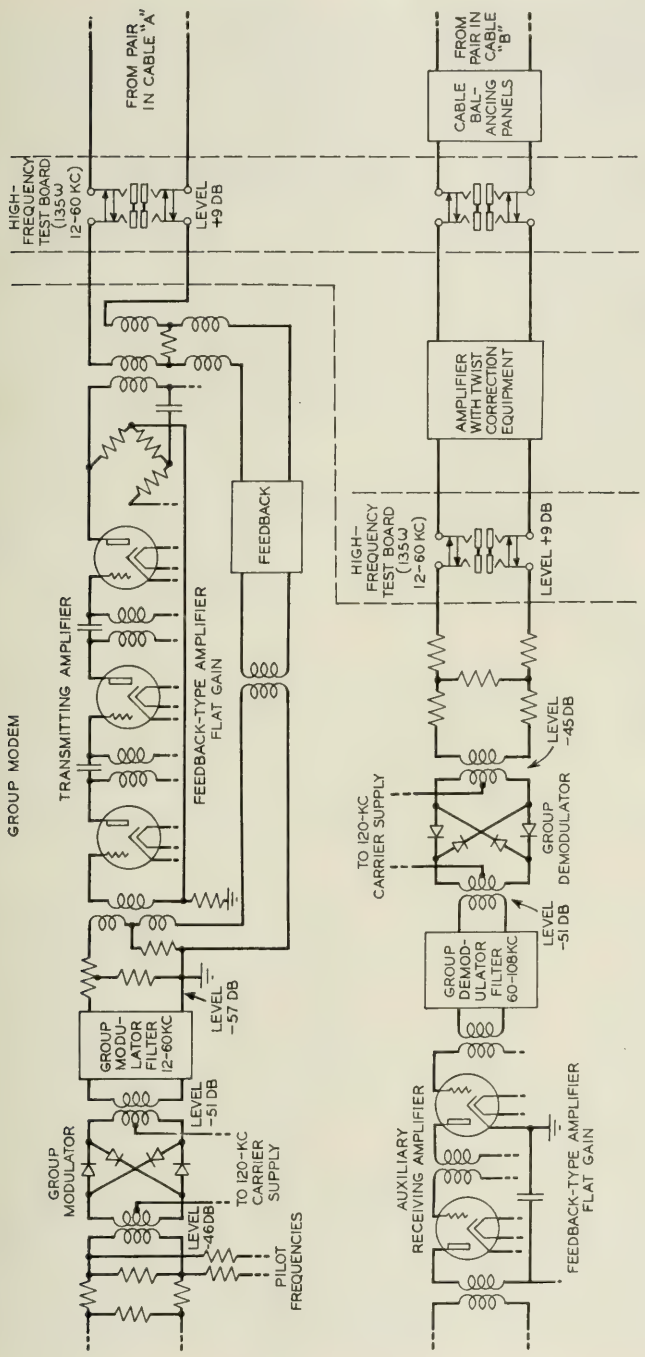


Fig. 3—*Continued from page 110.*

It seems to have an indefinite life. (Some have been on life tests as modulators for about five years.) The carrier power required is about 1/2 milliwatt to modulate satisfactorily a single telephone circuit level of − 13 db.

The modulator produces the usual two sidebands and the lower one is selected by the quartz crystal channel filter described in another paper.[4] This sideband, joined by eleven others, is stepped down to about the iterative impedance of the shielded office cabling. In the office cabling the twelve channels pass through the high-frequency patching bay to the double balanced group modulator of copper oxide where they are joined by three pilot channel frequencies.

The group modulator uses the same copper oxide as that in the channel modulator described above, but the carrier power is about 50 times greater (about 25 milliwatts) in order to keep down unwanted modulation produced between the twelve sidebands. To that same end the level of each sideband is made low (− 46 db), and the double balanced type of circuit is used to balance out some of the undesired products. It also balances out the twelve incoming bands in the range 60 to 108 kc. from the output and so simplifies the following group modulator filter.

From a level of − 57 db the twelve channels, now in the range from 12 to 60 kc., are amplified to + 9 db for delivery to the 19-gauge pair in the lead covered toll cable. The amplifier is a three-tube negative feedback type, using pentodes and operating with 154 volts plate battery which is composed of the usual 24-volt filament battery and 130-volt plate battery in series. The last tube is a power tube and does not overload until a single-frequency output of about one watt is reached.

On the receiving side in Fig. 3, the twelve incoming channels, in the range from 12 to 60 kc., pass from the amplifying and regulating equipment,[3] to the group demodulator. This is identical with the group modulator described above and transfers the twelve channels to the range 60 to 108 kc. The channels are then amplified to a − 5 db level by an amplifier of the negative feedback type using two low-power pentodes with 154-volt plate battery as described above for the transmitting amplifier.

From there the twelve channels are separated by the filters which are identical with those on the transmitting side, and are then demodulated and amplified to a + 4 db level as shown for one channel in Fig. 3. The demodulator is identical with the modulator but it is poled oppositely on the carrier supply so that the d-c. components of modulation in the modulator and demodulator neutralize each other

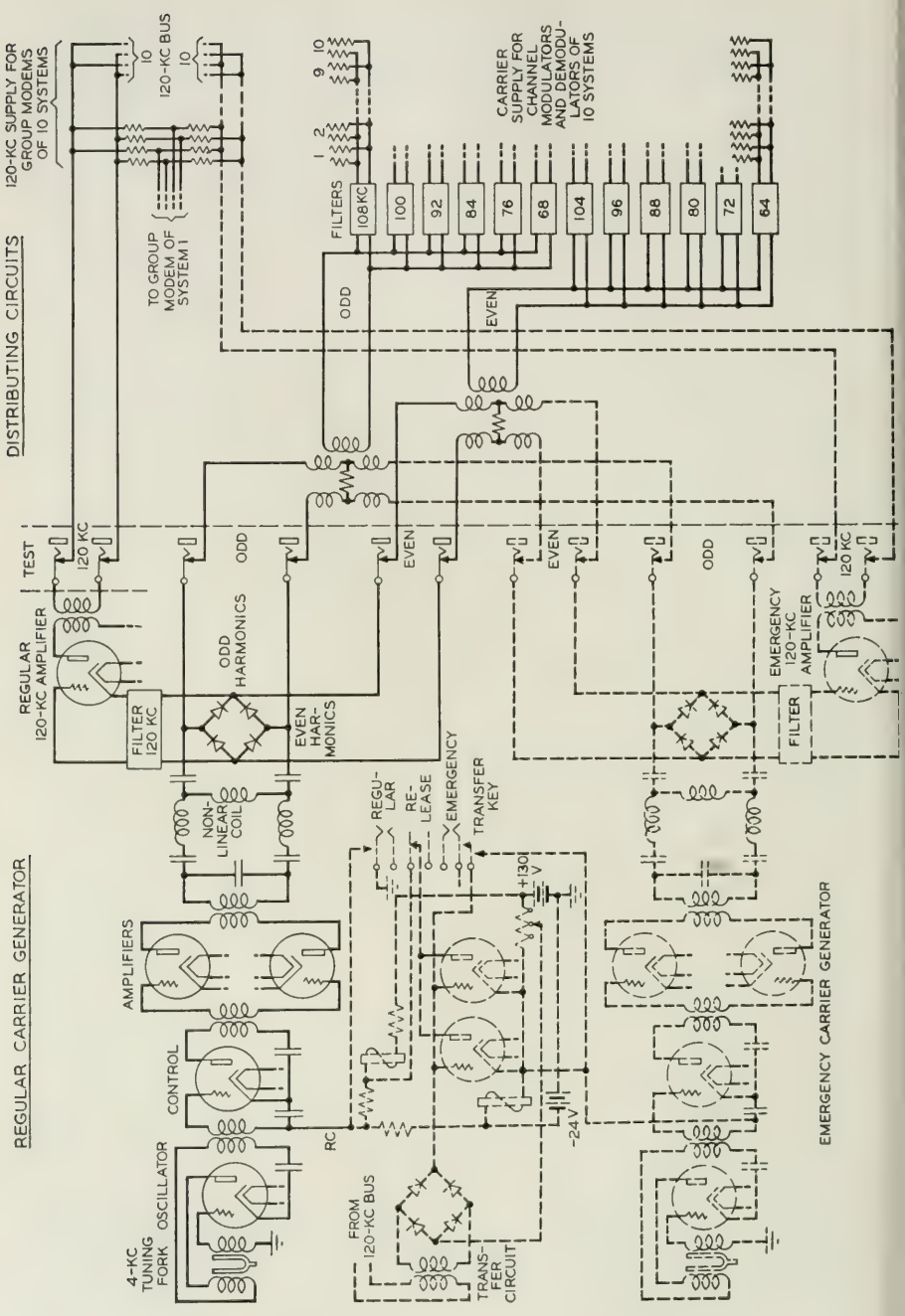
and thereby avoid developing an undesirable voltage bias. The poling also reduces somewhat the amount by which stray frequencies have to be suppressed in the carrier supply. The demodulator amplifier has a slide wire gain control rheostat to equalize channel levels, which functions by changing both the grid bias on the tube and the amount of negative feedback which is introduced by the rheostat. The sliding contact in the slide wire is made practically free from contact trouble by the space current of the tube flowing through it. As the rheostats are only about 1000 ohms and small in size, they can easily be mounted at a distance from the amplifier in the voice-frequency jack field.

The carrier supply for the twelve channels from 64 to 108 kc., and for the group modems of 120 kc., is derived in the circuit shown in Fig. 4. A regular generator is shown at the top in solid lines and an emergency generator at the bottom is shown in dotted lines. Between the two is an automatic transfer circuit (in dotted lines) which transfers to the emergency whenever the regular generator fails to supply the proper amount of 120 kc. to the 120 kc. bus.

At the upper left-hand corner is shown a 4-kc. tuning fork, of an alloy having a low temperature coefficient driven by the tube to its right to operate as an oscillator of very stable frequency. The next, or control tube, amplifies the 4 kc. to drive the push-pull power stage where a power of about 4 watts is developed. This passes through the 4 kc. filter to the non-linear coil where odd harmonics of 4 kc. are produced. The underlying principles of operation of this coil have been published.[5] To derive even harmonics of 4 kc., the copper-oxide bridge is used which rectifies about half the energy of the complex wave of odd harmonics but, by balance, greatly reduces the amount of the odd harmonics present in its output. Odd harmonics are obtained at one point and even at the other. This separation into odd and even harmonics by the balance of the copper-oxide bridge provides effective loss of about 30 to 40 db and reduces the requirements on the carrier supply filters which follow.

The two branches pass through hybrid coils to the banks of channel carrier supply filters. These separate the frequencies and feed them to twelve carrier supply bus bars, one for each channel frequency. From these the individual modems are fed through protective resistances so that an accidental short circuit on one of the modems will not cut off the carrier supply to the others.

The hybrid coils permit the two generators to be connected so that either can feed into the same bank of channel carrier supply filters without being reacted upon by the other. No switching is required when changing from regular to emergency supply.


120-KC SUPPLY FOR GROUP MODEMS OF 10 SYSTEMS
10
120-KC BUS
10
TO GROUP MODEM OF SYSTEM 1
DISTRIBUTING CIRCUITS
1 2 9 10
FILTERS
108 KC
100
92
84
76
68
CARRIER SUPPLY FOR CHANNEL MODULATORS AND DEMODULATORS OF 10 SYSTEMS
104
96
88
80
72
64
ODD
EVEN
TEST
120 KC
ODD
EVEN
EVEN
ODD
120 KC
REGULAR 120-KC AMPLIFIER
FILTER 120 KC
ODD HARMONICS
EVEN HARMONICS
EMERGENCY 120-KC AMPLIFIER
FILTER
NON-LINEAR COIL
REGU-LAR
RE-LEASE
EMERGENCY
TRANSFER KEY
+130 V
REGULAR CARRIER GENERATOR
AMPLIFIERS
CONTROL
RC
-24 V
EMERGENCY CARRIER GENERATOR
4-KC TUNING FORK OSCILLATOR
FROM 120-KC BUS
TRANS-FER CIRCUIT

The 120-kc. carrier which feeds the group modulators of ten systems or a total of one hundred and twenty talking channels must be very dependable. Therefore separate filters are used for the regular and emergency supply and separate amplifiers for the large power required by group modulators. Regular and emergency distributing buses are provided. Each group modulator and each group demodulator is wired through protective resistances to the regular bus and through another set of protective resistances to the emergency bus. With this arrangement an accidental short circuit even across one of the busses or across one of the output coils of one of the 120-kc. amplifiers will not stop the whole supply of 120 kilocycles.

The 4-kc. oscillator of the emergency generator is in constant operation so that when it is needed no time is required to start it, but the grid bias on the second tube is held above its cutoff value by the automatic transfer circuit. This prevents the 4 kc. from going further until called for in an emergency.

An emergency is indicated when there is no 120-kc. supply on either the regular or emergency bus. When this happens, the copper-oxide rectifier in the transfer circuit gets no 120 kc. and so loses its rectified voltage. This triggers off one or both of the two gas-filled tubes (multipled for safety) which increases the grid bias on the control tube of the regular generator to stop its 4 kc. supply and at the same instant restores the bias to normal on the control tube of the emergency to let its 4 kc. pass through and put the whole emergency circuit into operation. The keys in the transfer circuit are provided for maintenance purposes, and to return from emergency to regular operation, since the gas tube circuit is arranged to transfer automatically in only one direction.

The pilot supply circuit is shown in Fig. 5. The 3.9-kc. tuning fork oscillator at the left supplies that frequency, through the three transformers, to the three copper-oxide modulators the carriers of which are obtained from the regular channel carrier supply bus-bars as shown. The three filters, which are identical with channel carrier supply filters, select the lower sidebands to be used for pilot frequencies at 64.1, 92.1 and 104.1 kc. The three pilot frequencies are distributed to the different systems through protective resistances from a bus-bar as shown. They are set 100 cycles off the carrier frequencies to obtain locations of minimum interference from carrier leak and other sources.

Signaling circuits do not form an integral part of the carrier terminal equipment. Signaling equipment of a type already widely used in the Bell System for toll circuits, is connected between the toll switchboard and the four-wire terminating set of the individual channel.

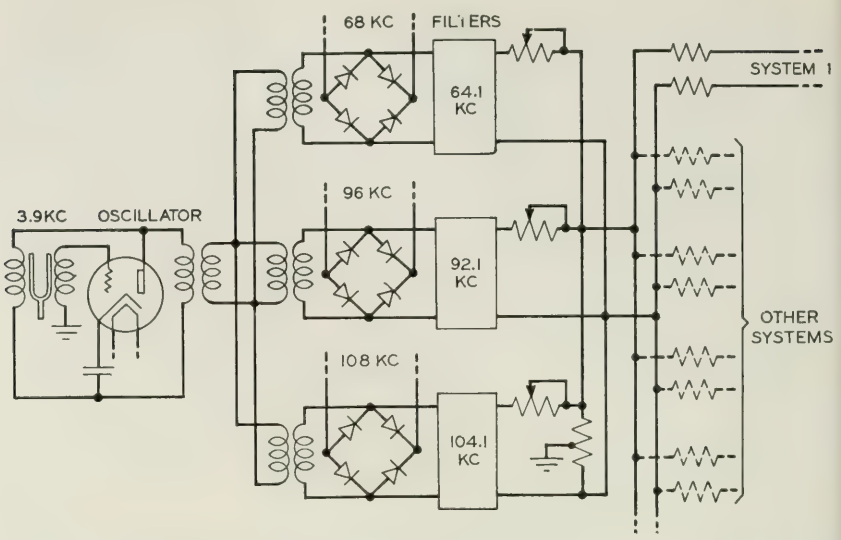


Fig. 5—Pilot supply circuit.

Transmission Performance

In general, the performance requirements set down as objectives in the development of this system were based on the assumption that five carrier links operating in tandem and over a 4000-mile circuit should give satisfactory, high-grade service.

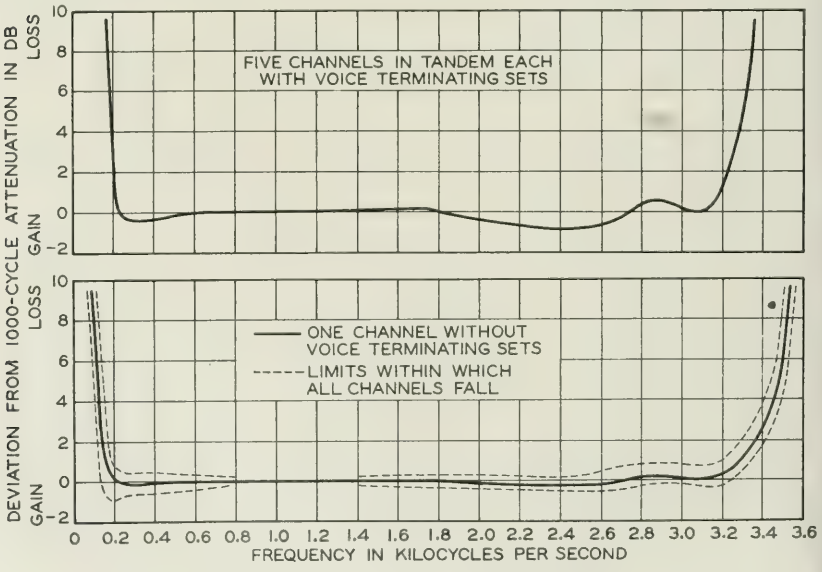

DEVIATION FROM 1000-CYCLE ATTENUATION IN DB
LOSS
GAIN
FIVE CHANNELS IN TANDEM EACH WITH VOICE TERMINATING SETS
ONE CHANNEL WITHOUT VOICE TERMINATING SETS
LIMITS WITHIN WHICH ALL CHANNELS FALL
FREQUENCY IN KILOCYCLES PER SECOND

Fig. 6—Channel frequency characteristics.

The channel frequency characteristic which has been attained in the terminals is shown in Fig. 6. The solid curve below shows the frequency characteristic of a representative channel, while the dotted curves near it show the limits within which the characteristics of all single channels, so far measured, would fall. Above in the figure is shown the characteristic of five representative channels in tandem, each channel having its two voice terminating circuits included.

The delay distortion and time of transmission, contributed by all terminal apparatus at both ends of a system except voice terminating sets, are shown in Fig. 7 for a single channel.

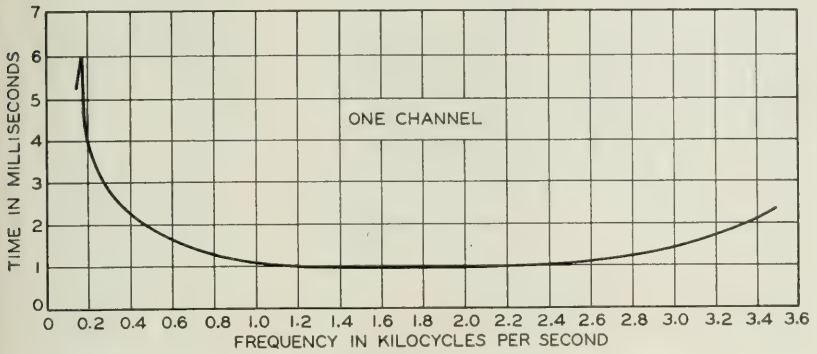


Fig. 7—Delay characteristic.

The channel modulators have been adjusted so that they will cut off the peaks of excessively loud talk to prevent overloading the carrier repeaters or other parts of the circuit, but this cutting is not enough to degrade the quality of speech. The single-frequency load curve of one complete channel is plotted in two ways in Fig. 8.

The frequency stability of the oscillating tuning forks is expected to be within $\pm 1 \times 10^{-6}$ parts per degree Fahrenheit on all systems, with negligible variations due to other causes. The amplitude stability of each frequency at its distributing bus is expected to be within $\pm$ 1/4 db over a period of months. The impedances of the bus-bars are sufficiently low so that crosstalk from one system into another through this path is unimportant. The effectiveness of the protective resistances at the carrier supply bus-bars is such that a short circuit on one modulator or demodulator will increase the loss in the remaining modulators and demodulators less than 1/2 db. The speed of switchover to emergency carrier supply is such that the disturbance to transmission will be less than 10 milliseconds. The effect on speech is not detectable.

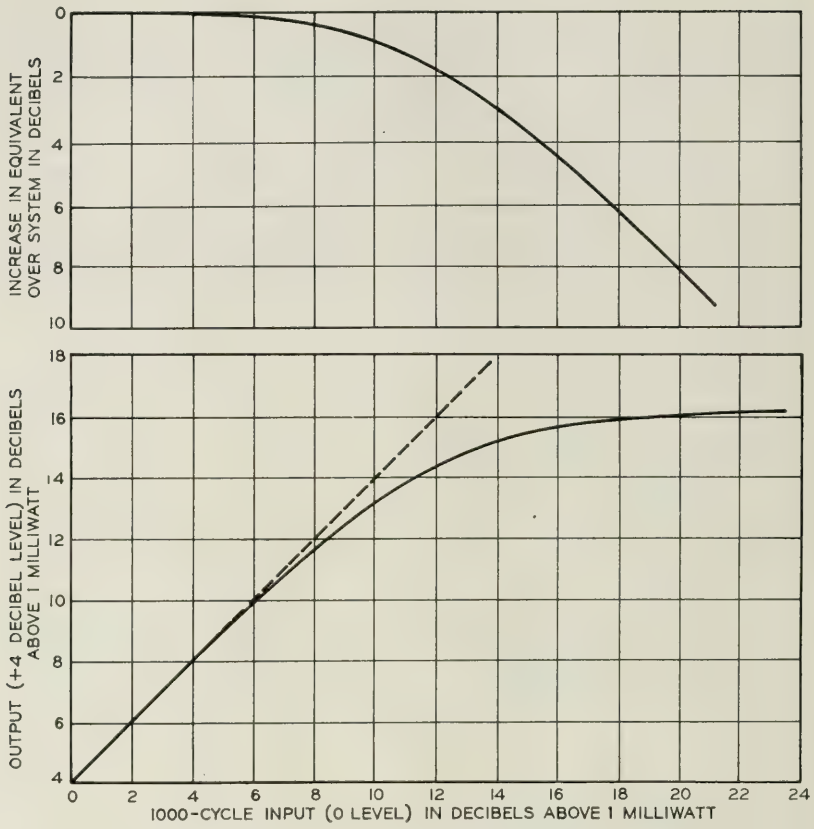


Fig. 8—Channel load curve at 1000 cycles.

Maintenance Features

Since the type "K" system provides more circuits in a group than ever before, it is essential that appropriately better maintenance facilities be furnished. Wherever vacuum tubes are used, jacks have been included to permit testing the condition of the tube by plugging in a new type of test set. The testing of a working tube with this set will not produce an appreciable reaction on performance of the circuits involved. When it has been determined that a tube in the common equipment is nearing the end of its useful life, a special transfer cord circuit is used to remove the circuit involving the tube from service and to substitute a spare circuit temporarily while the defective tube is replaced. This transfer from a regular to a spare and vice versa can be made without effect upon service.

In a type K terminal office transmission tests are made at the four-wire test board shown in Fig. 9, *A*, where the incoming and outgoing voice-frequency circuits appear, and at the high-frequency test board shown in Fig. 9, *B*. At the former, four-wire talking, monitoring and testing circuits have been provided for voice-frequency maintenance.

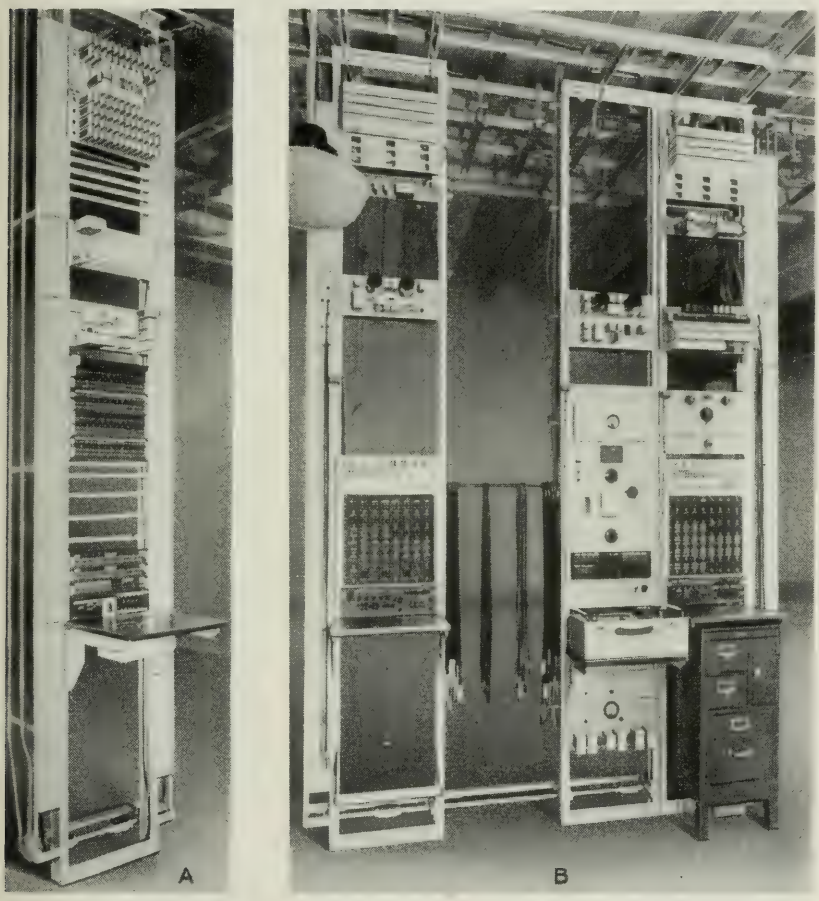


Fig. 9—Test positions: "A"—voice frequency; "B"—high frequency.

Adjustment of the equivalents of the individual channels can be made from this point as previously described. Much can be done from this position by means of monitoring and noise measuring to diagnose troubles.

At the high-frequency testboard the circuits are brought through jacks and high-frequency measuring apparatus is provided. Measure-

ments can be made on operating systems to determine the performance of the intermediate repeaters and regulators with respect to level and equalization over the frequency range from 12 to 60 kilocycles. Loss and gain measurements also can be made either between this point and the voice-frequency four-wire test board, through the carrier terminal equipment or through the next adjacent repeater or terminal office at high frequencies. It is possible to test the high-frequency portion of the terminal and to substitute a spare, by patching or rapid transfer, for a defective or potentially defective group modulator, transmitting amplifier, group demodulator or receiving amplifier.

Some of the high-frequency testing equipment is shown mounted on the middle bay of Fig. 9, *B*, of which one of the most important units is the 1 to 150-kc. test oscillator located at the center of this bay. It is a heterodyne type of oscillator which covers the frequency range with a continuous film strip scale about 300 inches long. Its maximum output is about one watt and this varies less than 1 db over the entire range. It is provided with built-in calibrating features and can be set to any frequency with an absolute accuracy of about 25 cycles. It is used as the tuning control of the pilot level measuring circuit. An auxiliary scale on the oscillator permits tuning the measuring circuit directly in terms of frequency.

The pilot level measuring circuit is of the double heterodyne type and includes a copper-oxide modulator which is supplied with carrier from the heterodyne oscillator, an intermediate frequency 130-kc. crystal filter of 10 cycles band width, a high-frequency amplifier for this frequency, a copper-oxide demodulator supplied with carrier from a 129-kc. fixed frequency oscillator, a voice-frequency amplifier and calibrating circuit. The input impedance of the measuring circuit is high so that when it is bridged across a line pair at the high-frequency testboard jack fields it does not produce appreciable loss to the line. The circuit permits measuring each of the three pilot frequencies to check levels and equalization of operating systems. The panels comprising the circuit are shown below the oscillator in Fig. 9, *B*.

Mounted on a shelf just below the oscillator is the transmission measuring set, which contains a highly accurate thermocouple and meter combination with calibrating circuits, wide range repeating coils, a test key circuit, and attenuators, one of which can be set in steps of 1 db up to a total of 90 db.

Equipment Features

Because of the large number of systems likely to be terminated in an office, the jacks are concentrated in a group of bays located together

for ease in patching and testing. There are in general five major divisions of the terminal equipment consisting of channel modem bays, group modem bays, carrier supply bays, high-frequency testboard, and four-wire voice-frequency patching board and associated voice termin-

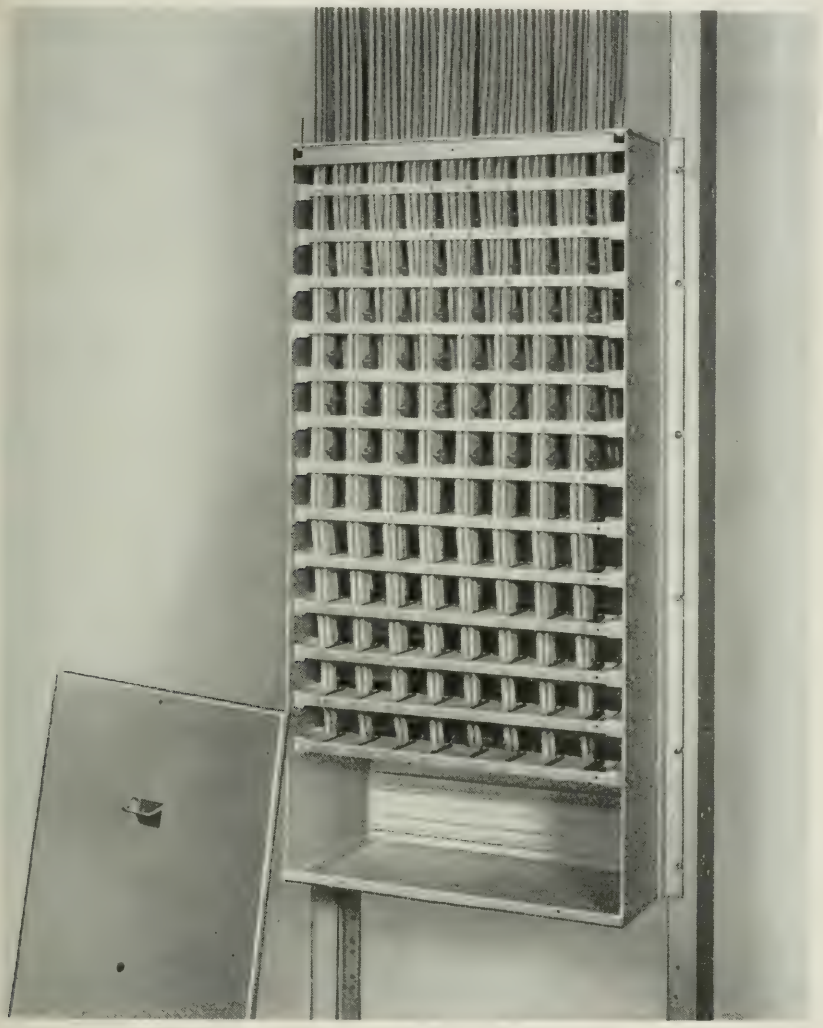


Fig. 10—Cabling of high-frequency jack bay.

ating equipment. The general arrangement in an office is such as to simplify the cabling between various groups of equipment

The cabling of a high-frequency jack bay, shown in Fig. 10, illustrates the congested wiring condition occurring when a large number of heavy

shielded wires is run to one location. Because of this congestion the jacks in this bay are mounted or removed from the front.

The concentration of equipment in the modem unit is made possible by the small size of the copper-oxide bridges and the filters. Fig. 11, *A*, shows twelve modem units for two systems on adjacent bays with space left at the bottom for the six modems of a third system.

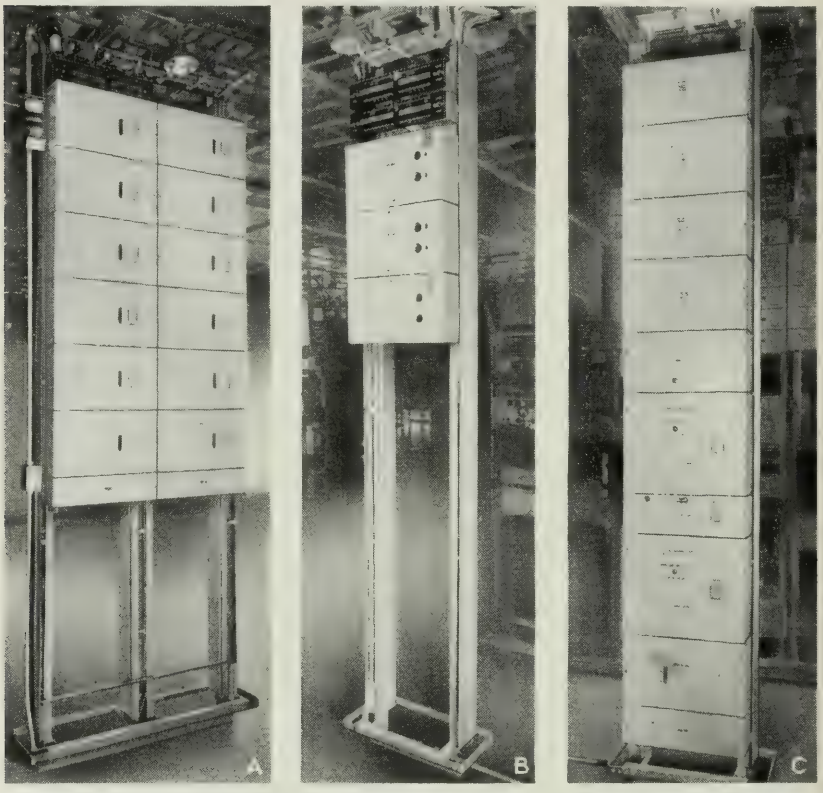


Fig. 11—Carrier equipment bays: "A"—channel; "B"—group; "C"—carrier and pilot supply.

The group modems are about the same size as the channel modems and include a modulator, a demodulator, a transmitting amplifier, and auxiliary receiving amplifier with associated filters. Fig. 11, *B*, shows three units for three systems with space for six additional units at the bottom of the bay.

The carrier supply equipment for ten systems is mounted in one bay as shown in Fig. 11, *C*, which includes the regular and emergency generators, transfer unit, distributing equipment and pilot channel supply

panel. The bays of this type are located near their associated channel equipment because the supply is chiefly for channel modems. One carrier distributing unit provides for the even and another for the odd harmonics. All terminals and bus bars of these units which are common to the ten systems are protected by insulating covers.

The four-wire voice-frequency jacks for all the systems in an office will ordinarily be grouped in associated bays, one of which is shown on Fig. 9, *A*. A bay will accommodate five systems as an average, that is, 60 voice circuits including the necessary pads and telephone set.

The high-frequency testboard is an arrangement of sealed test terminals, high-frequency patching jacks and high-frequency testing equipment mounted on bays as shown in Fig. 9, *B*. Only a few high-frequency patching jacks were required initially and these were therefore mounted above the sealed terminals. This arrangement of bays with the addition of a high-frequency patching bay at the right of each sealed terminal bay will accommodate 100 systems.

The carrier pairs are split off from the main toll cables at splices in the cable vault. The input circuits are carried thence in lead covered cable to the cable crosstalk balancing bays and thence to the input sealed terminal. The output pairs run directly from the output sealed test terminal to the splice in the cable vault. The remaining high-frequency wiring from rack to rack is shielded wire.

Conclusion

The carrier telephone terminals for the type "K" system which have been described are simpler, occupy less space and provide better transmission performance than multi-channel carrier terminals used previously in the Bell System. As part of a general development of broad-band transmission systems, it is very desirable to employ equipment which can be used in common with several systems. The 12-channel bay, much of the carrier supply and all of the voice-frequency terminating equipment of this type "K" system terminal will be used to form corresponding parts of the terminals for the 12-channel open-wire system and the coaxial system, both of which are under development. This not only has simplified the development work, but also will result in greater mass production of these common parts and provide desired uniformity of voice-frequency terminating levels and maintenance arrangements.

References

1. "Carrier Current Communication on Submarine Cables," H. W. Hitchcock, *Jour. of A. I. E. E.*, October 1926; *Bell Sys. Tech. Jour.*, October 1926.

2. "New Key West-Havana Carrier Telephone Cable," H. A. Affel, W. S. Gorton and R. W. Chesnut, *Bell Sys. Tech. Jour.*, April 1932.
3. "Carrier Telephone Systems for Toll Cables," C. W. Green and E. I. Green. Presented at the 1938 Winter Convention of the A. I. E. E. Published in this issue of the *Bell System Technical Journal.*
4. "Crystal Channel Filters for the Cable Carrier System," C. E. Lane. Presented at the 1938 Winter Convention of the A. I. E. E. Published in this issue of the *Bell System Technical Journal.*
5. "Magnetic Generation of a Group of Harmonics," E. Peterson, J. M. Manley and L. R. Wrathall, *Electrical Engineering*, August 1937.

Crystal Channel Filters for the Cable Carrier System * †

By C. E. LANE

SINCE the channel selecting filters used at the terminals of the twelve-channel cable carrier system are the principal filters in the system this paper is concerned primarily with these. Their importance is evident from the fact that they represent over one-third of the cost of the system terminals.

Many new features appear in these channel filters. The most outstanding is the use as filter elements, along with inductance coils and condensers, of plates cut from crystalline quartz. It is for this reason they are called "crystal filters." In addition, however, the inductance coils, some of the condensers, and also the filter assemblies have in them new features. Only after a number of years of laboratory experimentation with filters using crystal elements, studying their advantages and limitations, was the cable carrier system planned to use such filters.

There are twelve channel filters which transmit the lower side bands derived from the modulation of the speech signals with carrier frequencies spaced 4 kilocycles apart from 64 to 108 kilocycles. An insertion loss frequency characteristic which applies for each of the twelve filters is shown in Fig. 1. Regarding a 10 db loss increase as the cut-off as compared with transmission at 1000 cycles, the voice-frequency band for a single-carrier link, largely determined by the characteristics of the channel filters, extends from approximately 150 to 3600 cycles. For five links the band extends from about 200 to 3300 cycles. This is a 600 or 700 cycles wider frequency band than the present three-channel open-wire carrier system. The maximum delay distortion in the transmission band of each of the filters is about 0.4 millisecond. As many as ten of these filters may appear in tandem in the longest talking circuits. The total delay distortion in such cases would then not exceed 4 milliseconds. This is not objectionable since the average listener can not observe the effect of delay distortion unless it exceeds about 10 milliseconds. A representative filter schematic is shown in Fig. 2. The condenser shown by the dotted line at the left is used only in the two lowest frequency filters to obtain

* This is a companion paper to other papers covering different parts of the twelve-channel cable carrier system.

† Presented at Winter Convention of A.I.E.E., New York, N. Y., Jan. 24–28, 1938.

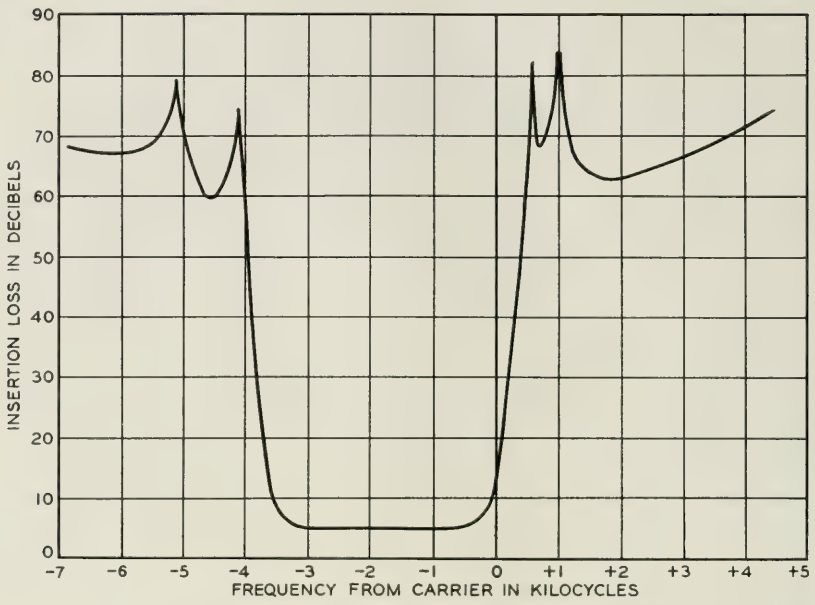


Fig. 1—When plotted in cycles removed from the carrier frequency, the insertion loss frequency characteristics of each of the twelve crystal channel filters are for all practical purposes identical.

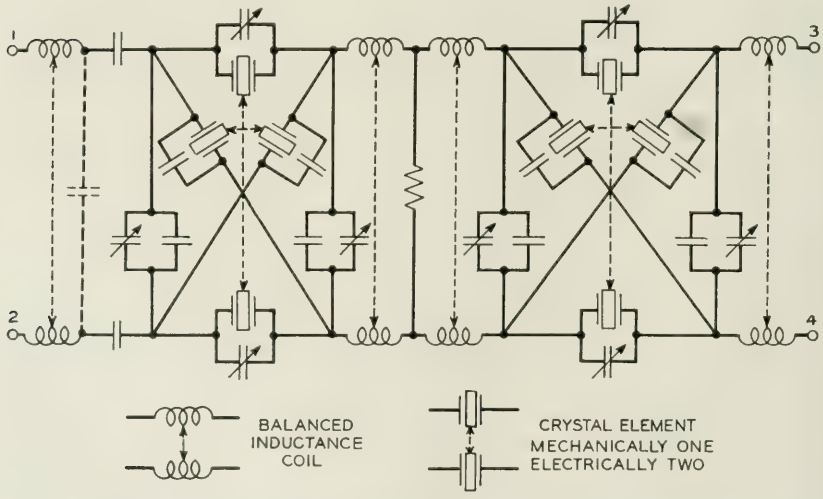


Fig. 2—The schematic circuits of each of the twelve filters are the same except for the addition of the condenser shown by the dotted lines which appears only in the two lowest frequency filters.

an impedance transformation internal to the filters and thereby permit the use of crystals of practical thicknesses for these filters. However, the equivalent circuit for each of the twelve filters is the same. In the system the filters work in parallel at one end and between terminating impedances of 600 ohms. The two condensers appearing in the series arms at the left end of the filter schematic are used in obtaining satisfactory operation of the filters in parallel and otherwise might be omitted provided the inductance at this end was made smaller at the same time.

Figure 2 indicates the separate physical elements and the manner in which these are connected in the filters. In considering the performance of the filters the crystal elements are replaced by their equivalents, an inductance and capacitance in series, shunted by a second capacitance as shown in Fig. 3. Also, the condensers in

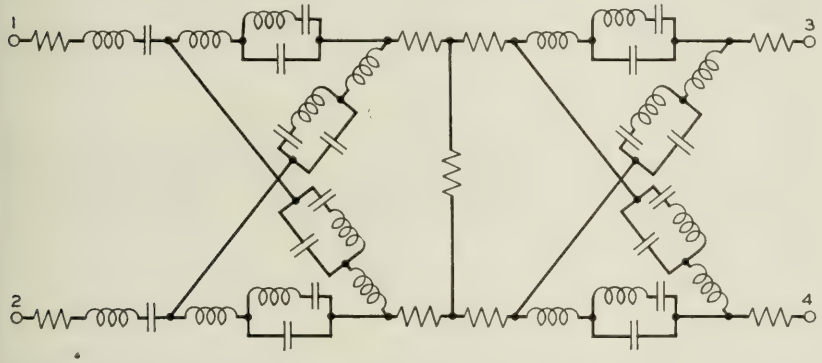


Fig. 3—The schematic circuits of the filters are each equivalent to two lattice sections with a resistance pad between them, resistances at each end and at the paralleling end a coil and condenser which resonate at the mid-band frequency.

shunt across the filter are shown inside the lattice combined with the direct capacitance of the crystals and the inductances are relocated in series in each lattice arm. In making this conversion, however, the effective resistance of the inductance coils are, for reasons which will appear later, shown remaining outside the lattice. Also the capacitances and the portion of the inductance which are used solely for purpose of paralleling are left outside the lattice. The basis for the conversion from Fig. 2 to Fig. 3 is shown in Fig. 4.

Before considering further the filter as a whole, the nature of the crystal elements and the reason for using them will be considered. It is common knowledge to those familiar with the performance of electrical wave filters that the energy loss unavoidably associated

with inductances imposes limitations upon the filter characteristics obtainable. Capacitances may be designed so that the energy dissipation is small and negligible as compared to that in the inductances. With ideal reactance elements entirely free from dissipation, filters might be designed for any band width with as little loss in the band as wanted and at the same time frequencies might be rejected outside the band by any amount desired, no matter how near such frequencies

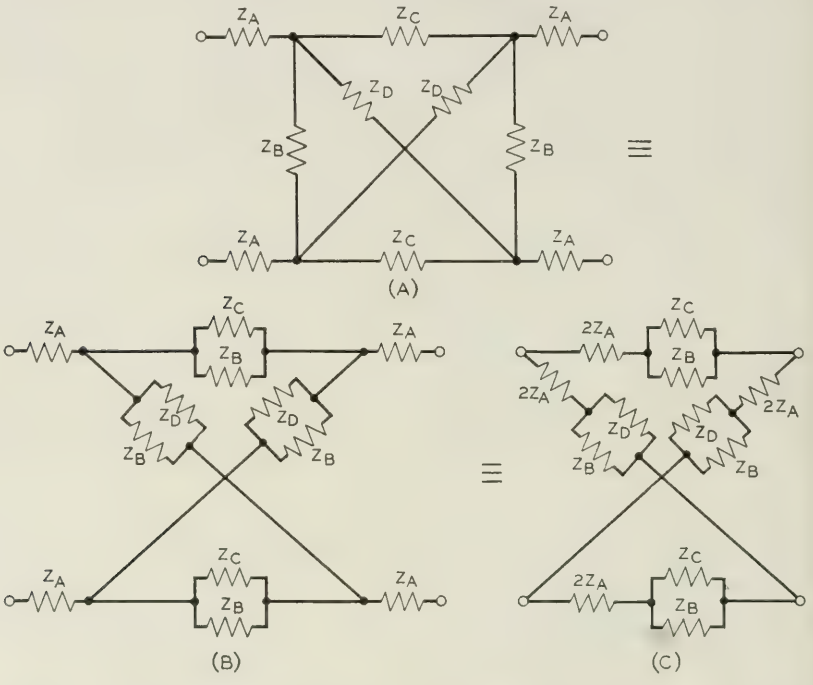


Fig. 4—It does not alter the transmission properties of a network such as shown in Figure 4A to remove the impedances in shunt and outside the lattice and replace them by impedances of equal magnitude in shunt across each lattice arm nor by removing the impedances in series with the lattice and replacing them by series impedances inside the lattice of twice the magnitude of those removed.

were to the edges of the transmitting band. Of course the sharper the filter cut-offs, other requirements being the same, the more complex the filter structure would be even neglecting dissipation. The greater the dissipation in the filter elements, the greater the loss in the transmitting range of the filters and the greater the number of cycles required for this loss to rise from the relatively low and uniform loss in the transmitting band to the high loss wanted outside the band. In the design of channel filters for carrier systems, the presence of

dissipation in the filter elements is costly in that the channels must be spaced farther apart than would otherwise be necessary, thereby wasting frequency space. At the same time the loss to transmitted frequencies must be made up for by amplification. The amount of dissipation in a reactance element is measured by the ratio of the effective resistance component of its impedance to the reactance component at any frequency. The reciprocal of this ratio is called the Q of the reactance and hence is a measure of efficiency or freedom from dissipation. In the design of inductances in the form of wire wound coils, it is generally not practical to obtain Q's much in excess of 200 or 300 at any frequency. The quartz crystal element used in the filters as previously stated is equivalent electrically to a two-terminal reactance consisting of an inductance and capacitance in series shunted by a second capacitance. For the Q of the inductance in the equivalent circuit of the crystal element a value of 15,000 or more can readily be obtained. It is for the purpose of utilizing this high Q inductance and obtaining the benefits therefrom that crystal elements are used in these filters.

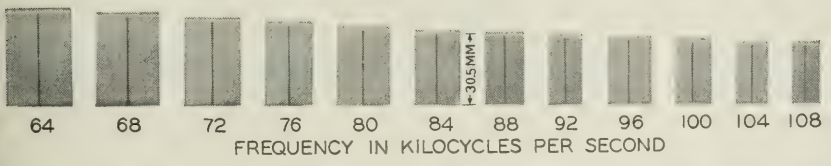


Fig. 5—The length of the crystal elements used in the different filters varies about inversely as the frequency of the filter band location.

The filter schematic in Fig. 2 shows crystal elements in each filter section; the two in the lattice arms and the two in the series arms in each case are identical. Electrically there are four crystals in each section but for reasons of economy and for convenience in handling and adjusting the crystals those in corresponding arms are physically one. This is possible since the filter is a balanced structure and the two like crystals vibrate in unison. Figure 5 is a photograph which shows a representative double crystal element taken from each of the twelve filters. The four crystals in the lowest frequency filter range from 40.2 millimeters to 41.8 millimeters in length and those in the highest frequency filter from 23.8 to 24.3 millimeters. The thickness of the crystals in all four of the lowest frequency filters are 0.63 milimeters, in the next four filters 0.82 millimeters, and in the highest frequency filters 1.1 millimeters. Uniformity in thickness is maintained as far as practicable since it contributes to economy in manu-

facture of the crystal. Within the range using the same crystal thickness the impedance and frequency differences, called for by the design of the different filters, can be provided by variations in width and length of the crystals. The ratio of width of the crystals to their length ranges from about 1/2 to 4/5.

The major surfaces of the crystals are plated with a thin layer of aluminum deposited by an evaporation process. This plating is divided along the center line lengthwise of the crystals to form the two

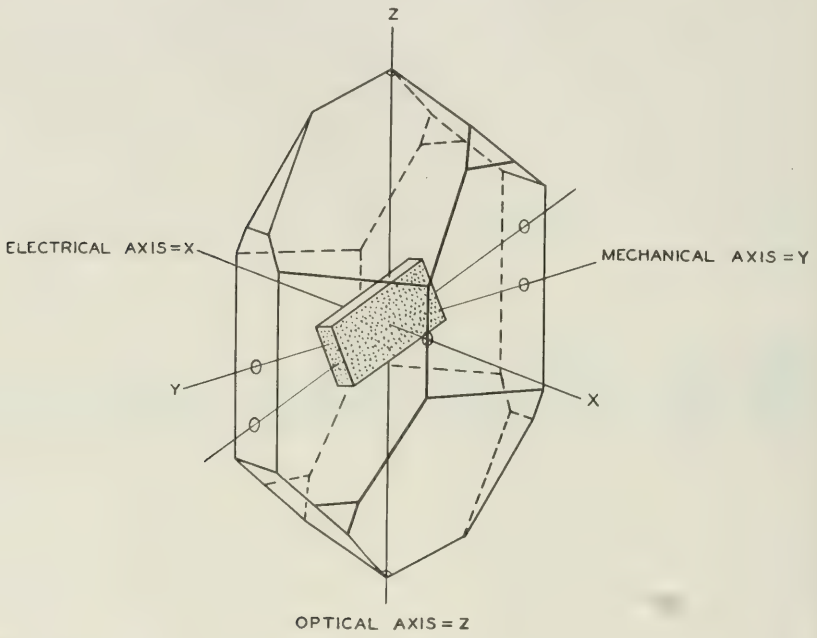


Fig. 6—The crystal elements are out with their major surfaces perpendicular to the electrical axis of the natural quartz, with their side surfaces making a small angle with the mechanical axis, and with their end surfaces making a small angle with the optical axis.

electrically independent crystals. Since the crystals vibrate longitudinally with a node across the middle, they are clamped at this node in mounting. Figure 6 shows the orientation of the crystal plates with respect to the natural axes of the quartz from which they are cut. The plates are cut as accurately as is practicable to the dimensions computed making a small allowance in length and then the crystal is finally adjusted in an electric circuit by grinding the end of the crystal until the resonant frequency falls within five or ten cycles of that desired.

Considering again the filter schematic as a whole (Fig. 3) and neglecting the dissipation in the crystals and condensers, the filter may be regarded as made of two lattice filter sections having ideal reactance elements, that is, elements free from dissipation. The location of the effective resistance of the coils outside the lattice, for purpose of performance analysis, shows how at the end of the filter these resistances may be regarded as part of the terminating impedance between which

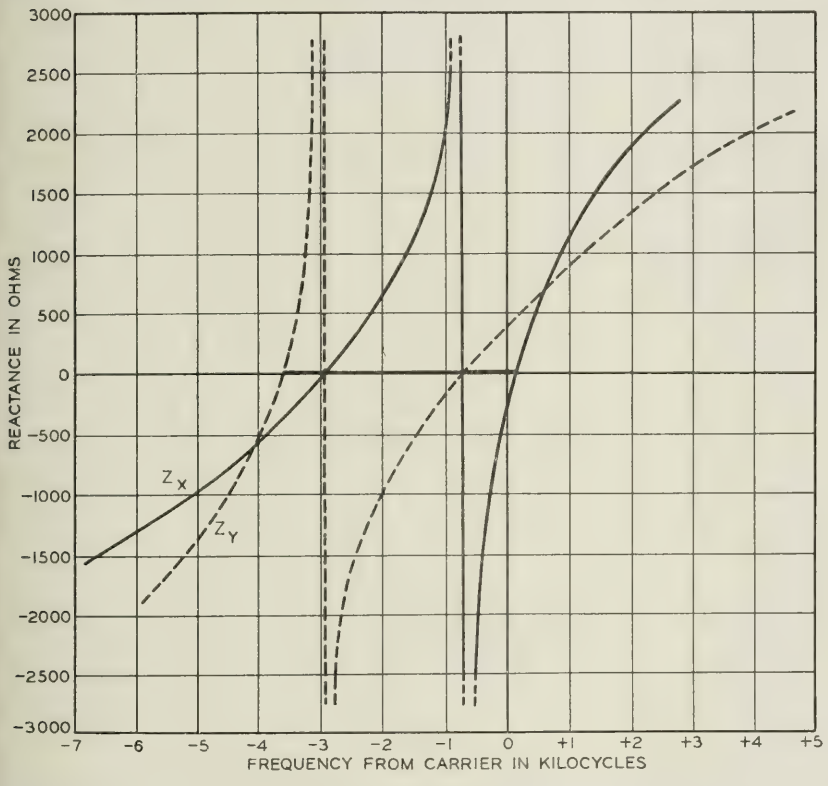


Fig. 7—In a lattice filter section transmission occurs for frequencies where the impedances for the two arms of the lattice are of opposite sign and attenuation peaks of very high loss occur where the impedances cross.

the filter works, and how between the filter sections the resistances may be combined with a shunt resistance to form a resistance pad which matches the image impedance of the two filter sections. The effect, then, of the coil resistances is primarily to provide a flat loss over the entire frequency range and does not affect appreciably the shape of the loss characteristic furnished by the reactance inside the lattice sections.

In considering the performance of lattice type filter sections, it is common practice to sketch together the frequency reactance curve of the two lattice arms Z_x and Z_y. This is done for one of the filter sections and is shown in Fig. 7. In the frequency range where the two curves are of opposite sign the filter transmits, and where they are of

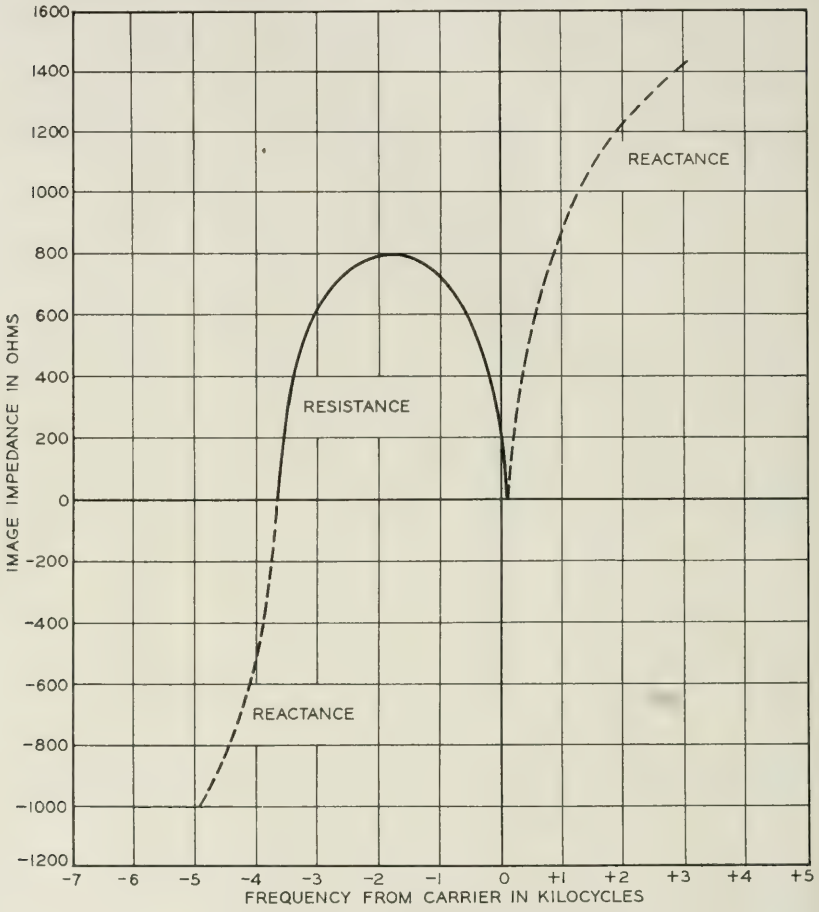


Fig. 8—The large reflection losses occurring within the transmission band of the filters and near the edges of this band has the effect of narrowing the width of the transmission band.

the same sign there is attenuation. At the point where the two curves intersect there are attenuation peaks of very high loss. The reactance curves of Fig. 7 are for the filter section accountable for the pair of attenuation peaks shown in the filter characteristic which are the closer to the edges of the transmitted band. For the other section the cross-

over points of the two reactance curves are farther away from the band, since this section is responsible for the outer pair of attenuation peaks. The design of the filters consisted in determining values for the inductance coils, condensers, and crystals, such that the reactance curves of the lattice arms of the filter passed through infinity and intersected with each other at the desired frequencies and, at the same time, in determining the impedance level for all of the elements such that the filters would have the right image impedances.

The curves of Fig. 7 would seem to indicate a somewhat greater band width for these filters than shown by the insertion loss characteristic of Fig. 1. The reason for this can best be explained by referring to the image impedance of one of the filter sections as shown in Fig. 8. Within the band the image impedance is, of course, a pure resistance which varies with frequency. It is about 800 ohms at mid-band frequency and falls rapidly to zero near the edges of the band. Assuming the effective resistance of the coils, which is about 100 ohms, as belonging to the terminating impedances, the filter sections actually work between impedances of about 700 ohms. This means that large reflection losses occur at each end of each filter section near the edges of the transmission band where the image impedance of the filter is very small. It is these reflection losses that are responsible for the actual transmission band being much narrower than it would be with the filter sections terminated in their actual image impedances. The filter sections are designed with 800 ohms image impedance at mid-band frequency instead of 700 ohms to make the band flatter and somewhat wider than it would be otherwise.

When a number of band filters are operated in parallel it is generally necessary to connect across the paralleled end a two-terminal network to correct for the distortion that would otherwise be present in the highest- and lowest-frequency filters in the group. A circuit of the network used for this purpose with the channel filters is shown in Fig. 9.

The filters employ crystal elements in order to obtain abrupt discrimination between wanted and unwanted frequencies and at the same time to secure low and uniform loss in their transmitting bands. This characteristic must not only be obtained at the time the filters are assembled and adjusted but must be maintained throughout the service life of the filters and not appreciably affected by temperature variations. This imposes severe stability requirements upon the elements used in the filters. The crystal elements themselves are very stable when properly designed and once adjusted will retain at a given temperature their frequencies of resonance within one or two cycles seemingly indefinitely. Their temperature coefficient is only

about twenty-five parts per million per degree centigrade, which is not objectionable.

The obtaining of inductance coils and condensers that were adequate in stability for use in conjunction with the crystals required considerable development effort. The inductance coils were required to have not only a high degree of stability with respect to temperature and time but also a high ratio of reactance to effective resistance, low modulation, and at the same time be small in size. Air core coils might have been designed for the purpose but they would have been quite large. The coils used are of the toroidal type wound on about one and three-fourths inch rings of molybdenum-permalloy. To reduce eddy current losses the cores are made of very fine powder and then annealed to reduce hysteresis losses. The particles are mixed with insulating

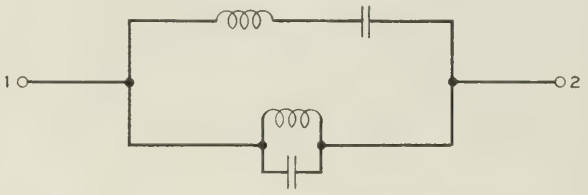


Fig. 9—A two-terminal reactance network is connected in shunt across the filters at their paralleling end to improve the characteristic of the highest and lowest frequency filters.

material and formed into rings by extremely high pressure. The inductance of the coils has a temperature coefficient of less than 40 parts per million per degree centigrade. The cores of the coils for the higher-frequency filters are wound with finely stranded wire to help secure good Q's (about 225). Because of the high impedance of the coils called for by the filter design, care is taken to make the capacitance between the windings and the core and between the windings and the case as low as practicable and also to make stable all such small capacitances as must be present.

The two extra condensers used at one end of each filter for paralleling purposes are of a high grade mica type. The other condensers are all quite special. The fixed ones, ranging in magnitude from about 7 mmf to 100 mmf, are made by plating short lengths of high grade glass tubing inside and outside with silver. Because of the intimate association of electrodes with the surfaces of the tubes and the low expansion coefficient of the glass used, a condenser is obtained that has a temperature coefficient comparable with that of the coils and crystals. No aging effect has been observed. It will be noticed that four small

adjustable air condensers appear in each filter section. These are used to secure precise initial adjustment of the filter capacitances.

To protect the filter elements against moisture, the filters are hermetically sealed in a container made from a rectangular section of seamless brass tubing with closely fitting plates soldered in each end. One end plate carries four metal-glass seal terminals and a nozzle through which dry air is blown after the filter is assembled. The other end plate is provided with a small hole for the escape of the drying air. After the drying operation the hole in the nozzle and the hole in the opposite end of the filter are closed by soldering. The elements that make up the filter are assembled on a chassis which is completely wired and then slid into the container in assembly. Figure 10 is a photograph showing this chassis and the arrangement of the

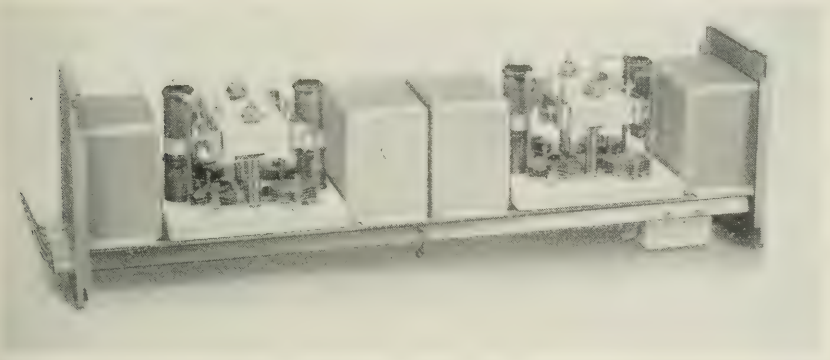


Fig. 10—The filter parts are assembled and wired on a chassis which is slid as a unit inside the filter container.

elements. The elements are located in such a way as to use very short wiring connections which reduce the magnitude of any stray admittances. The wired filter chassis is carefully adjusted by setting the values of the air condenser such that for each filter section the resonance frequencies looking into each end of each section occur where they theoretically should. This compensates for the effect of small capacitances between the filter parts. In the design of the filter parts care is taken to use no material which absorbs moisture readily since such moisture would later be released and raise the relative humidity of the air inside the filter.

If a potential much in excess of about twenty volts is applied across crystal elements at frequencies near resonance, the crystals will break from the mechanical strain of their vibration. The maximum safe voltage across the channel filters at the resonant frequencies of the

crystals is considerably greater, however, since at resonance the full voltage does not appear across the crystals. In normal use the voltages across the filters will be very much less than twenty volts.

Other filters forming part of the terminal apparatus are the group modulator and group demodulator low-pass filters, the channel and group carrier supply filters, and the pilot supply filters. The group modulator and demodulator filters are of the low-pass type employing only coils and condensers as elements. The group carrier supply filter is the same in schematic and mechanical design as the crystal channel filters described. The pilot supply filters and the channel carrier supply filters are equivalent in schematic to one section of the channel filters; but of course they are only about half the size and are hermetically sealed in the same manner.

References

For a further discussion of crystal filters the reader is referred to "The Evolution of the Crystal Wave Filter" by O. E. Buckley, *Journal of Applied Physics*, October 1936, and "Electrical Wave Filters Employing Quartz Crystals as Elements" by W. P. Mason, *Bell System Technical Journal*, July 1934.

Crosstalk and Noise Features of Cable Carrier Telephone System *

By M. A. WEAVER, R. S. TUCKER and P. S. DARNELL

CROSSTALK and noise are important factors in cable carrier transmission as outlined in the paper "A Carrier Telephone System for Toll Cables" by Messrs. C. W. Green and E. I. Green. Crosstalk and noise limit the number of carrier channels which can be utilized in any one cable, not only by limiting the number of channels which can be placed on a single pair, but by limiting the number of pairs which can be used. Noise also controls the transmission loss which can be permitted between repeaters. Without the crosstalk and noise reduction measures described in this paper, the number of carrier channels per cable would be so few and the spacing between repeaters so short, that the type K carrier system would be impracticable.

Crosstalk

To utilize existing toll cables in the Bell System for frequencies up to 60 kilycycles required the solution of many new crosstalk problems because: (1) Crosstalk increases rapidly with the frequency, (2) Non-loaded carrier pairs due to their high speed of propagation are especially suitable for very long distances and hence the crosstalk requirements per unit length are relatively severe, (3) The large gains of the carrier repeaters amplify certain crosstalk currents much more than in the case of voice frequency circuits.

Two general effects need to be considered: intelligible crosstalk must be prevented; and, a large number of circuits crosstalking into a particular circuit must not contribute an undue amount of noise. The second effect is called babble, since it consists of a multiplicity of unrelated voice sounds which, in the aggregate, are unintelligible.

An important feature is the use of different cables for opposite directions of transmission. This makes the major crosstalk problem the reduction of crosstalk between pairs in the same cable used for transmission in the same direction. The crosstalk currents due to transmission at one end of a disturbing circuit through the distributed couplings with a disturbed circuit tend to arrive at the distant end at the same time since the currents via any of the couplings travel sub-

* Presented at Winter Convention of A. I. E. E., Jan. 24–28, 1938.

stantially the same distance. This makes it possible to greatly reduce the total effect of these distributed couplings by the use of small adjustable mutual inductance coils connected between pairs at one point in each repeater section.

If nothing more were done, there would still be objectionable crosstalk since currents from the outputs of carrier repeaters could crosstalk into voice frequency circuits and these circuits could then again crosstalk into other carrier frequency circuits at points near their repeater inputs. This effect is minimized by transposing the carrier pairs from one cable to the other at carrier repeater points.

At common voice frequency and carrier frequency repeater points there would be an unsatisfactory crosstalk path from a carrier repeater output into all the wires not used for carrier frequencies and from them through coupling between office wiring into similar wires in the other cable and finally into carrier repeater inputs in the second cable. This crosstalk is minimized by the use of carrier frequency suppression coils in the voice frequency circuits. These coils also serve the purpose of preventing carrier frequency noise originating in voice frequency circuits from being transmitted into the cables and inducing noise at points near carrier repeater inputs.

Near-End Crosstalk

Near-end crosstalk is the result of coupling between circuits transmitting in opposite directions, while far-end crosstalk is the result of coupling between circuits transmitting in the same direction. Near-end crosstalk coupling between different carrier circuits of the same frequency must be kept very small, particularly near a repeater point, since crosstalk from the output of a repeater into an opposite directional pair near the input of its repeater will be greatly amplified by this repeater.

Crosstalk between carrier circuits within the offices is kept low by careful shielding, segregation, suppression of spurious paths through battery supply, common grounding arrangements, etc.

Since the type K system operates on a "four-wire" basis, different electrical paths are used for opposite directions of transmission. Satisfactory near-end coupling in the outside plant is obtained, therefore, by placing east bound pairs in one cable and west bound pairs in another. When two cables have relatively heavy sheaths as in the larger Bell System cables, their coupling is sufficiently small even with the two cables in close proximity.

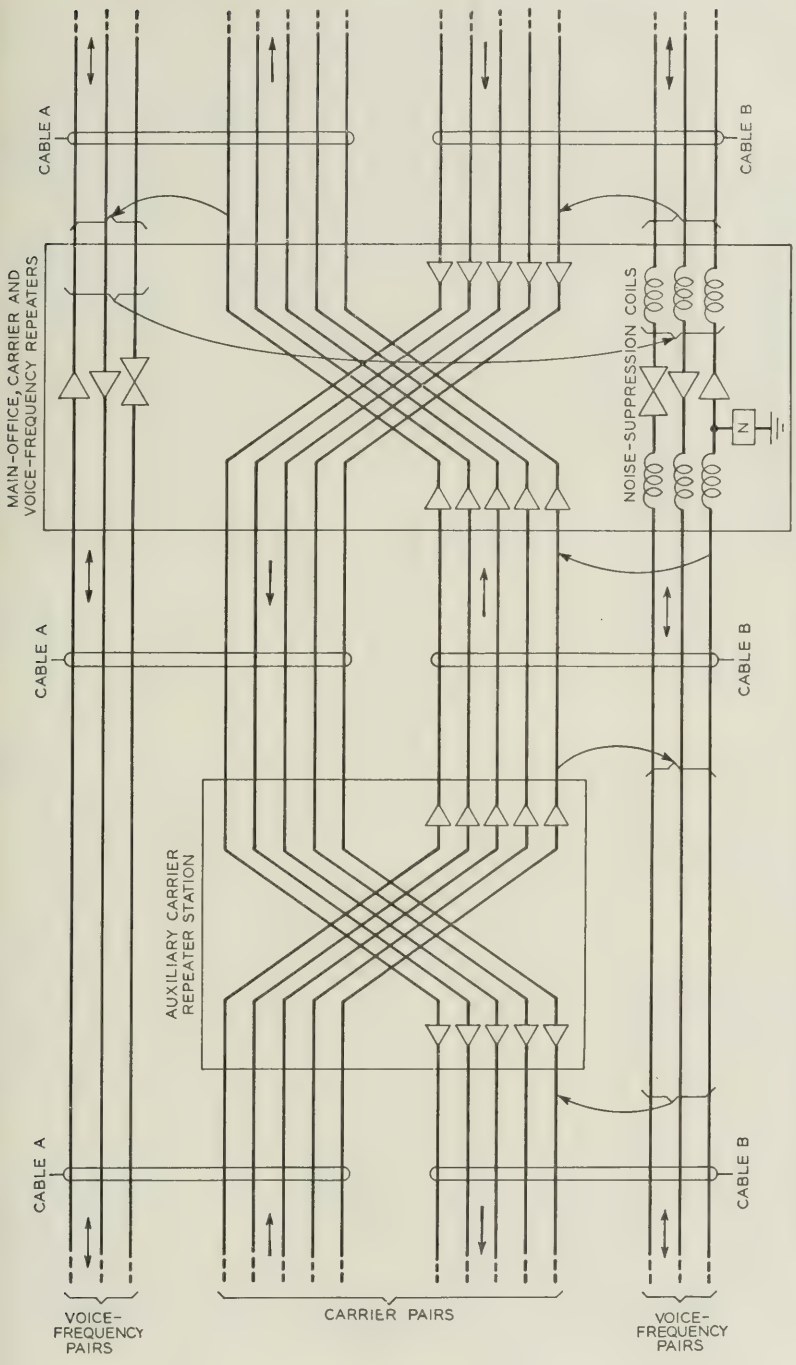


Fig. 1—Schematic showing carrier pairs transposed between cables, coupling paths and noise suppression coils.

Interaction Crosstalk

The crosstalk currents from a carrier repeater output into voice frequency circuits in the same cable must be limited, since they crosstalk again into carrier circuits near repeater inputs and, consequently, are amplified by the high gain repeaters. Intermediate circuits most responsible for crosstalk of this type are made up of combinations of pairs and phantoms and the sheath, i.e., longitudinal paths.

One case of crosstalk of this kind would occur if the same cable were used for carrier pairs transmitting in the same direction on both sides of a repeater. This is prevented by transposing carrier pairs from one cable to the other at each repeater point, as shown on Fig. 1.

A second interaction crosstalk problem is encountered at the common voice and carrier repeater points and involves coupling between cables as well as in the same cable. Here the coupling path is from carrier repeater outputs to intermediate circuits in the same outside cable, back into the common office over these intermediate circuits and then via office coupling to intermediate circuits in a second outside cable and from there to carrier repeater inputs connected to pairs in the second cable. Referring to Fig. 1, a set of noise (and crosstalk) suppression coils is encountered in this path. The high longitudinal circuit impedance of these coils minimizes this interaction crosstalk.

Far-End Crosstalk

Far-end crosstalk currents are subjected to line attenuation and amplification similarly to the main transmission currents, and do not have extra amplification as in the case of near-end crosstalk. Furthermore, far-end crosstalk currents due to couplings at different points along the line tend to arrive at the distant end of the disturbed circuit at the same time. Hence a considerable portion of the far-end crosstalk over the type K frequency range, which occurs between circuits transmitting in the same direction in the same cable, can be neutralized by introducing compensating unbalances at only a comparatively few points, such as one per repeater section. The far-end crosstalk reduction problem is greatly simplified because phantom circuits are not used for carrier operation.

Theoretically, for the same precision of match between the impedances in the two directions at the balancing point, the crosstalk reduction would be about the same whether the balancing is done at an intermediate point or at either end of a repeater section. Balancing will be done at repeater inputs rather than at an intermediate point, such as the middle, because it is practicable to obtain repeater im-

pedances matching the average line impedance sufficiently well so that the effectiveness of balancing is reduced only slightly.

Nature of Far-End Crosstalk Coupling

The coupling between two cable pairs in a short length may be represented by a mutual admittance and a mutual impedance. The former is due almost entirely to capacitance unbalance, which varies but little with frequency, so that its effect could be practically balanced out by means of a simple condenser. The latter, however, involves a complex mutual inductance of the form $M_a + jM_b$, because of the proximity effect of the wires of a pair and of other cable conductors.[1] As shown on Fig. 2, both components vary considerably with fre-

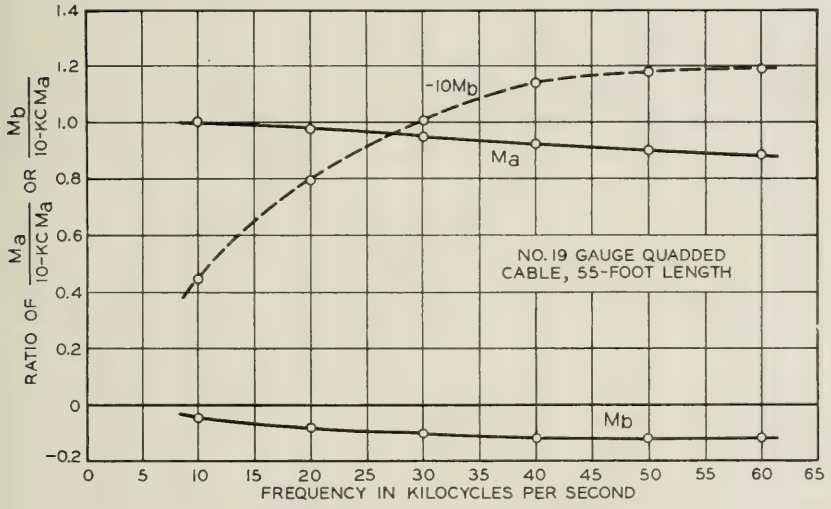


Fig. 2—Mutual inductance between cable pairs in terms of value for M_a at 10 kilocycles.

quency; M_a on the average decreasing as the frequency increases while M_b in the general case is of negative sign and reaches a maximum value at 56 kilocycles.

Type of Balancing

To obtain maximum reduction in crosstalk it would be necessary to use a condenser for balancing the mutual admittance and an inductance coil for balancing the mutual impedance or to use some equivalent complex network. Experimental balancing in a particular cable using the coil-condenser method reduced the mean crosstalk over the type K

[1] "Cable Crosstalk—Effect of Non-Uniform Current Distribution in the Wires," R. N. Hunter and R. P. Booth, *Bell System Technical Journal*, April 1935.

range about 20 db, which is close to the maximum reduction possible with a universal type of balancing unit. The reduction is limited by the fact that two pairs having identical crosstalk couplings in each of two short lengths at different points in the cable will not produce two identical elements of crosstalk current at a circuit terminal because: (1) Cable circuits are not perfectly smooth. Reflections, as at junctions of reel lengths or at terminals, alter the two crosstalk currents differently, (2) The propagation constants of each circuit vary slightly from reel to reel in a random fashion and therefore the two crosstalk currents are of slightly different phase and magnitude, (3) In any short length the disturbing circuit produces crosstalk currents in intermediate circuits, which are propagated along these circuits and crosstalk again into the disturbed circuit at various points, producing an additional crosstalk current at the circuit terminal. At any frequency, this interaction crosstalk current has a random phase and magnitude relation to the crosstalk current for the short length considered by itself, and depends also upon the position in a repeater section of the short length.

A 20 db crosstalk reduction is not required, considering the number of K systems anticipated in any one cable. Studies were made, therefore, to determine whether satisfactory results could be obtained with a less expensive type of balancing, as outlined below.

The effects of frequency and circuit impedance on crosstalk coupling are as follows: (1) Crosstalk in a short length due to capacitance and to inductance coupling increases about directly as the frequency increases for circuits whose impedance is independent of frequency. (2) Crosstalk due to capacitance coupling varies directly as the impedance of the circuits while that due to inductance coupling varies inversely as the impedance. Changing the impedance from about 800 ohms for loaded voice circuits to about 135 ohms for non-loaded carrier circuits and changing the frequency from about 4 kc. to 60 kc. increases the crosstalk due to capacitance coupling by a factor of about 2.5 and that due to inductance coupling by a factor of about 90.

Capacitive coupling in existing cables was reduced by design to as great a degree as practicable, particularly for the most closely associated circuits, because it is of most importance in the loaded voice frequency case. These same design measures also reduce inductive coupling but not to the same extent. Capacitive coupling decreases rapidly with separation due to the shielding effect of copper in intervening circuits while inductive coupling is not much affected by intervening copper wires. To minimize magnetic coupling it is necessary to use different lengths of twist for the pairs. Existing cables have relatively few lengths of pair twists.

As the net result, capacitance coupling is no longer all important, inductance coupling at 60 kc. actually predominating by a factor of about 3 to 1 in existing cables. Capacitance balancing should, therefore, be less effective than balancing designed to reduce the inductance coupling. Tests have shown that capacitance balancing alone gives a crosstalk reduction of about 11 db while inductance balancing alone gives a reduction of about 16 db. Since the latter reduction is sufficient, except possibly for small cables or special cases, the type K balancing has been designed on this basis. Far-end crosstalk currents due to the two kinds of coupling have phase relations not differing from zero or 180 degrees by more than about 15 to 40 degrees, depending on whether the upper or lower type K frequencies are considered. There is, therefore, a tendency for either type of balancing unit to annul both kinds of coupling.

To obtain as much as 16 db reduction it is necessary that the frequency characteristic of the balancing coil simulate that of the cable (Figure 2). This was found practicable, as discussed later, by shunting the primary (or secondary) of the coil by a properly designed impedance.

Size of Balancing Coil

To meet the crosstalk requirement it is necessary to balance each carrier pair against every other carrier pair. If 50 carrier pairs were used, there would be 49 balancing coils connected to each pair for balancing to all the other pairs, a total of 1225 coils. For convenience, adjustable coils having the same mutual inductance range and the same self-inductance are used. Hence, the insertion loss per coil, resulting from the self-inductance and resistance of the coils, must be kept small. In addition, the self-inductance of the coil presents a problem from the impedance standpoint. To keep the impedance at any point in the balancing panel as nearly like the average cable impedance as practicable, the self-inductance of a series of coils must be neutralized by capacitances shunted at suitable intervals. It is very desirable, therefore, to use coils whose self and mutual inductances are no larger than actually essential. Consequently, an attempt has been made to keep the maximum crosstalk before balancing low.

Due to special measures, described below, it appeared that the maximum inductance unbalance per repeater section could be kept below about 1.3 to 1.5 microhenries, with the possible exception of side-to-side unbalances, and trial balancing coils were designed accordingly.

Crosstalk Reduction Before Balancing

Changes in the original splicing are made at approximately 6000-foot intervals, i.e., at points where voice frequency loading coils must be

removed from the carrier pairs. In most existing voice frequency toll cables the 19-gauge quads were spliced as three groups, one a two-wire circuit group, one an east bound four-wire circuit group and the third a west bound four-wire circuit group. Ordinarily, the carrier pairs will be selected from the four-wire groups because these groups are usually larger than the two-wire group and since the quads within a group are spliced at random there is less chance of a large value of coupling between pairs of different quads, i.e., two pairs are less apt to be recurrently in a relation of high coupling. The carrier pairs are divided equally between the two four-wire groups, in order that the least number of four-wire voice circuits will be lost.

In cables with large four-wire groups it is satisfactory to maintain the grouping arrangement on the pairs converted to carrier. In such cables, however, one four-wire group is in the center or core of the cable and the other group in the outer periphery. In order that all circuits will have about the same velocity and attenuation and be subjected to about the same temperature conditions for both transmission and crosstalk reasons, one (four-wire) carrier group in these cables will be spliced to the other (four-wire) carrier group and vice versa at the 6000-foot intervals.

In cables with relatively small four-wire groups, there is more chance of two pairs being recurrently in a relation of high coupling. To reduce this chance, a special splicing arrangement has been devised for use at the 6000-foot intervals. With existing splicing the maximum coupling in cables with small groups is about 2.5 times that for cables with large groups. This ratio is appreciably reduced by the special splicing, likewise reducing the maximum mutual inductance that must be supplied by the balancing unit.

The foregoing was with particular reference to crosstalk between pairs in different quads. Crosstalk between pairs in the same quad (side-to-side crosstalk) is an additional problem. A quad consists of two twisted pairs of wires which are twisted together to permit the use of voice frequency phantom circuits. Since the two sides of a quad are so closely associated, side-to-side crosstalk is generally much greater that than between pairs of different quads. The electrical size of the balancing unit, therefore, is determined by the side-to-side crosstalk, which is reduced by "poling."

To apply poling, the quads are carried through as quads for an entire carrier repeater section. From measurements of side-to-side crosstalk in phase and magnitude, quads in one half repeater section are chosen and spliced to quads in the other half in such manner as to partially neutralize the side-to-side crosstalk. In effect, quads in one half-section serve as balancing units for the other half.

In most existing toll cables the side-to-side capacitance coupling was reduced when the cables were installed, by means of test-splicing within the 6000-foot sections. Obviously, for poling to be effective it is necessary to operate mainly on the inductance component. The poling measurements, therefore, are made at about 1 kc. where an approximate measure of the inductance component can be obtained directly since the capacitance and inductance components of the crosstalk are at an angle of almost 90° at this frequency. Figure 3 shows the crosstalk results obtained by means of 1-kc. poling on 14 repeater sections. It has been shown that this 9 db reduction is within 2 to 3 db of the

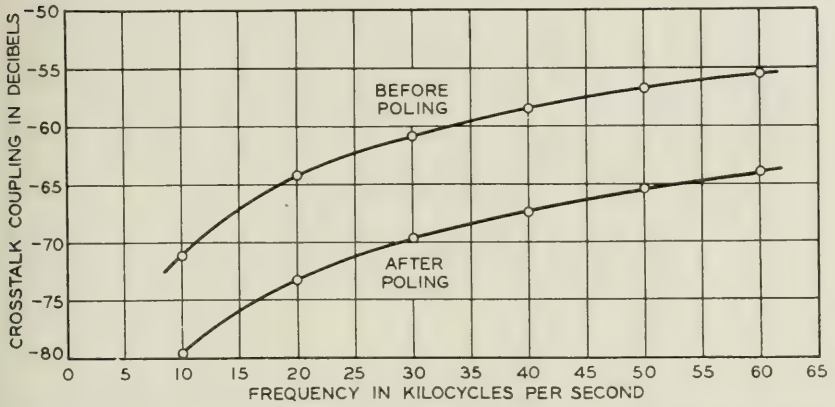


Fig. 3—R.M.S. side-to-side far-end crosstalk per repeater section from measurements on 14 repeater sections.

maximum reduction possible with much more complicated poling involving measurement and consideration of both components at carrier frequencies.

After side-to-side poling, coil balancing cannot be expected to give as much as 16 db reduction in crosstalk. This is unimportant, however, as long as the required reduction can be obtained more economically by the combined methods rather than by balancing alone.

Crosstalk Balancing Coil

Since the voltage which causes the crosstalk current in the disturbed circuit may be in either a clockwise or counter-clockwise direction, the balancing device, for flexibility reasons, should be capable of establishing voltages in either direction. A balancing coil was developed, therefore, which in operation may be likened to that of two separate transformers with simultaneously movable cores. The primary wind-

ings are in series, as are the secondary windings, and are connected as shown in Fig. 4, for example. The relative direction of each secondary winding is the same, whereas the relative directions of the primary windings are reversed. With the cores in mid-position, the voltages induced in the two secondaries are equal in magnitude but opposite in phase, and the net induced voltage in the disturbed circuit is zero. As the cores are moved toward the left the respective components of the voltage induced in circuit 3–7–8–4 increase in a counter-clockwise direction and decrease in a clockwise direction, the net result being a voltage in a counter-clockwise direction. Such a setting of the balanc-

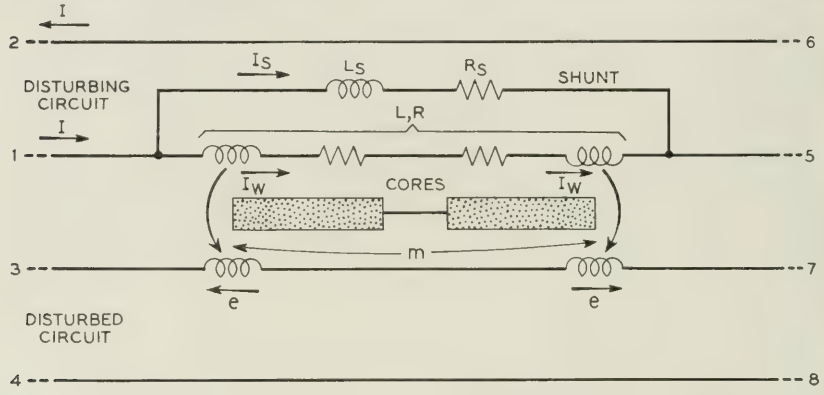


Fig. 4—Schematic of a simple balancing coil designed to produce a complex mutual impedance.

ing coil would be used to counteract a clockwise crosstalk voltage, the amount of departure of the cores from mid-position being dependent on the magnitude of the crosstalk voltage being counteracted. Movement of the cores toward the right produces the opposite effect

This device, disregarding any proximity effects therein and the effects of the shunt, acts to set up a net voltage e which is in phase quadrature with the disturbing current I. Hence,

$$e = -j\omega mI, \tag{1}$$

in which m is the net mutual inductance of the device. To obtain the required mutual impedance characteristic, the primary (or secondary) windings of the coil are shunted by an inductive resistance. Let the effective self-inductance and resistance of the line windings (primaries) be denoted by L and R, respectively, and the current through these windings by I_w. Let the effective self-inductance and resistance of the shunt be denoted by L_s and R_s, respectively. At balance, that is,

when no crosstalk current flows in 3–7–8–4 due to I (the disturbing current), the current I_w is

$$I_w = \left[\frac{R_s(R + R_s) + \omega^2 L_s(L + L_s)}{(R + R_s)^2 + \omega^2(L + L_s)^2} - j \frac{\omega(R_s L - R L_s)}{(R + R_s)^2 + \omega^2(L + L_s)^2} \right] I \quad (2)$$

$$= [a - jb]I, \quad (3)$$

where a and b are, respectively, the coefficients of the real and imaginary parts of the expression. Hence, with a shunted coil the voltage induced in the disturbed circuit is:

$$e = -j\omega m I_w = -j\omega(ma - jmb)I. \quad (4)$$

The mutual impedance, Z_m, equals $j\omega(ma - jmb)$, or the effective mutual inductance M of the balancing coil may be written

$$M = M_a + jM_b, \quad (5)$$

wherein $M_a = ma$ and $M_b = -mb$. Assuming R, L, L_s and R_s to be constant with respect to frequency of current and position of the cores, it is seen from (2) and (5) that for any core setting, M_a and M_b are functions of frequency only and their ratio at a given frequency is theoretically constant throughout the operating range.

To keep inductance L constant irrespective of the mutual inductance settings, the length of the coil windings, the length of the magnetic cores and their spacing with respect to the winding spacing are so related that the change in inductance of one primary (or secondary)

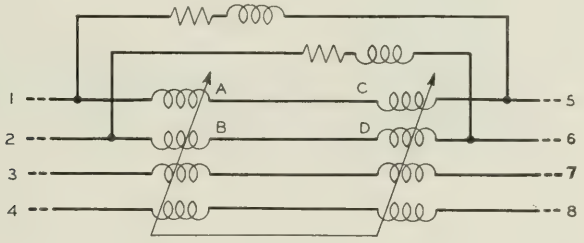


Fig. 5—Schematic of winding arrangement of trial balancing coil.

winding caused by motion of its associated core is equal and opposite to the change caused by the movement of the core associated with the other primary (or secondary) winding. To keep R low over the type K frequency range, cores of finely powdered molybdenum permalloy pressed into a cylindrical form are used.

Because of other requirements which a balancing coil must satisfy, the winding arrangement actually employed is shown in Fig. 5. The simple device of Fig. 4 is not balanced from the standpoint of longitudinal crosstalk for any coil setting except that of zero mutual induc-

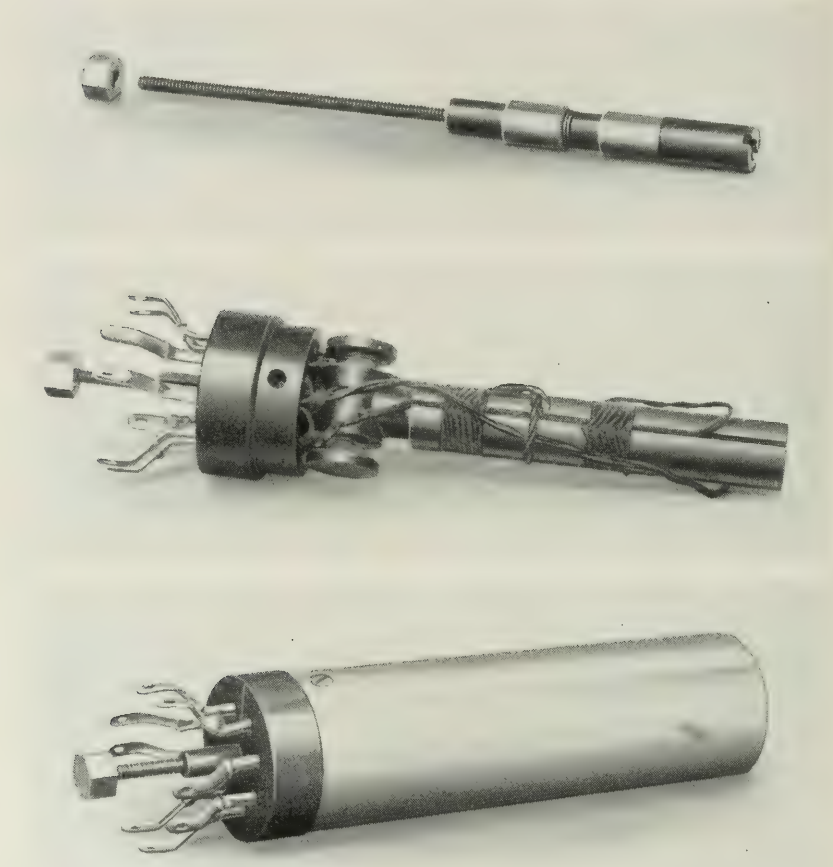


Fig. 6—Trial balancing coil construction. Container is 4½″ in length and 1⅜″ in diameter.

tance. The Fig. 5 arrangement is such that theoretically there is no magnetic coupling between the two circuits for longitudinal currents in either one, regardless of the coil setting. Unless the capacitance between primary and secondary windings can be kept very small, the resultant admittance unbalance produces crosstalk which is not com-

pletely balanced out when the coil is adjusted. The turns of conductor in the Fig. 5 coil are so located that this side-to-side capacitance unbalance is less than 5 micro-microfarads. The capacitances between wires of either the primary or secondary winding do not affect the unbalance but contribute a part of the capacitance loading which compensates for the line inductance of the coils.

In the actual balancing coil, shown in Fig. 6, the windings are located in channels cut in a fibre tube which is secured to a head carrying the winding terminals and a bushing through which passes the threaded brass rod supporting the two cores. Below the head are small spool

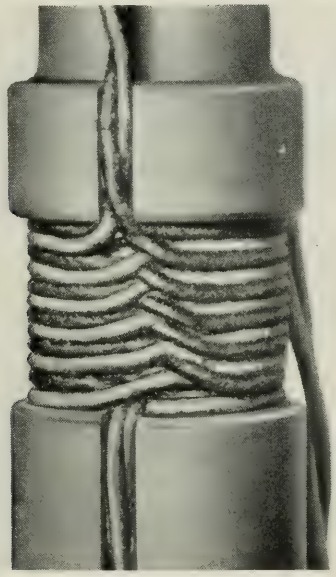


Fig. 7—Arrangement of inner winding of trial balancing coil.

forms on which the shunts are wound. Insulating material such as bakelite is used to obtain proper spacing of the two cores.

The rather unusual manner in which the turns are applied is illustrated in Fig. 7, which is a closeup view of the two wires forming the inner winding. These two wires alternately cross over each other, progressing along the axis in opposite directions of rotation. The outer winding is similarly applied. This type of winding eliminates all splices within the coil, removing hazards incident to interior splices.

The complete coil assembly is enclosed by an aluminum container which serves the dual purpose of a shield and a convenient means of holding the coil for mounting purposes as this container fits snugly into

an aluminum cup riveted to the assembly panel. The windings are dried and impregnated and the space between the coil assembly and container is filled with insulating compound.

The mutual inductance of a typical coil varies as shown in Fig. 8 as the cores are moved. The range, with the shunts disconnected, is approximately + 1.6 to − 1.6 microhenries, which is covered in about 16 turns of the screw (a total core travel of 0.5 inch). With the shunts connected, the effective mutual inductance at a given setting becomes less as the frequency rises, the two components, M_a and M_b, varying with frequency as shown in Fig. 9. To determine the proper values of L_s and R_s for the shunt, allowance must be made for the complex mutual inductance inherent in the coil due to proximity effect within the windings.

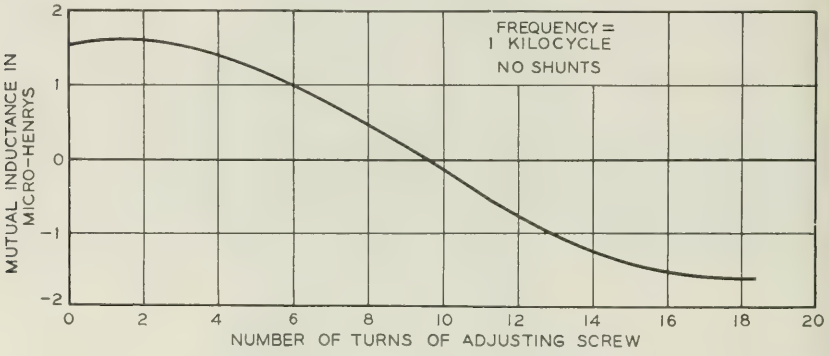


Fig. 8—Mutual inductance of trial balancing coil.

The series inductance of the balancing coil without shunts varies as shown in Fig. 10. As the cores are moved, the inductances of windings 1–*A*–*B*–2 and 5–*C*–*D*–6 (Fig. 5) behave as shown by their respective curves, one increasing as the other decreases. The sum of these two curves is shown by the dotted line, and the measured value of 1–5–6–2 is shown by the solid line. It is seen that the overall self-inductance of 1–5–6–2 is constant to within ± 0.1 microhenry. The difference between the curves (about 0.1 microhenry) is caused by the slight mutual inductance existing between winding 1–*A*–*B*–2 and 5–*C*–*D*–6, which is negative owing to reversed winding direction in this side of the balancing coil. The measured inductance around 3–7–8–4 would slightly exceed that obtained by adding the inductances of the two sections owing to positive mutual inductance between the two ends. These end effects could be reduced by greater separation of the two sets of windings, but this refinement is not necessary.

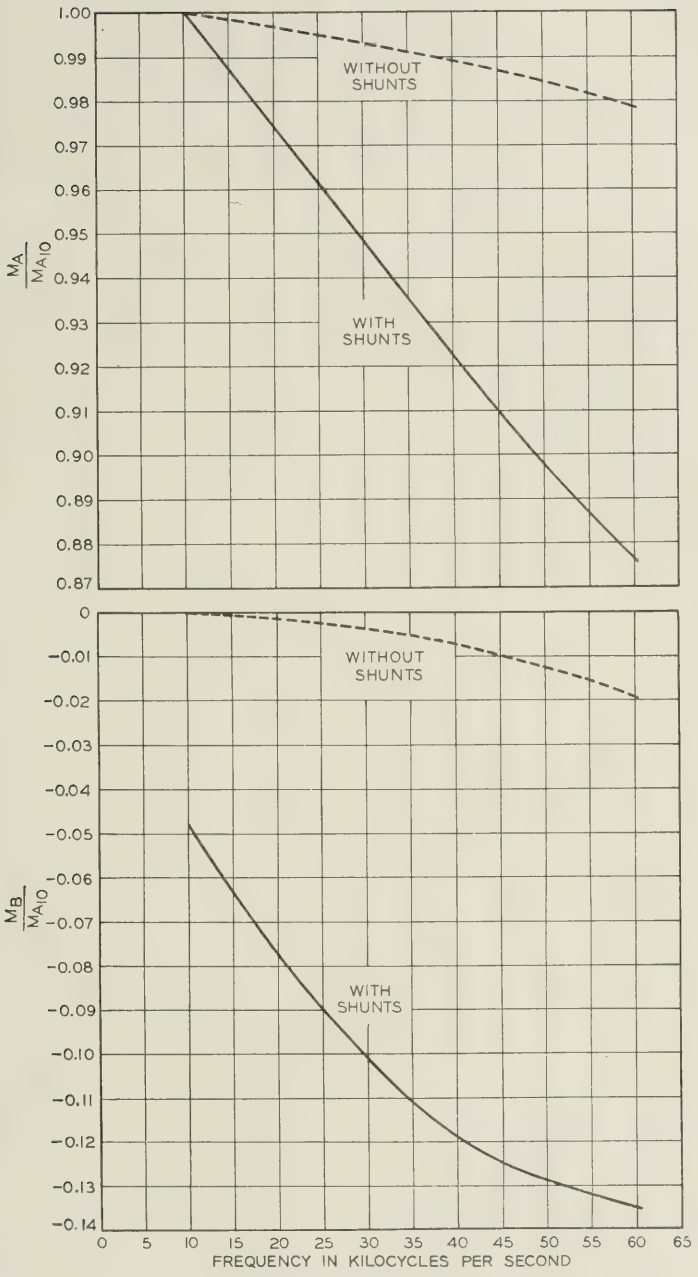


Fig. 9—Variation of M_a and M_b components of trial balancing coil with frequency.

When the shunts are connected, the inductance around 1–5–6–2 is lowered slightly, and the effective resistance is increased. To simplify the capacitance loading and in order not to introduce more resistance in one cable pair than another, the balancing coil assembly is so arranged that shunted and non-shunted windings are alternately introduced into a pair.

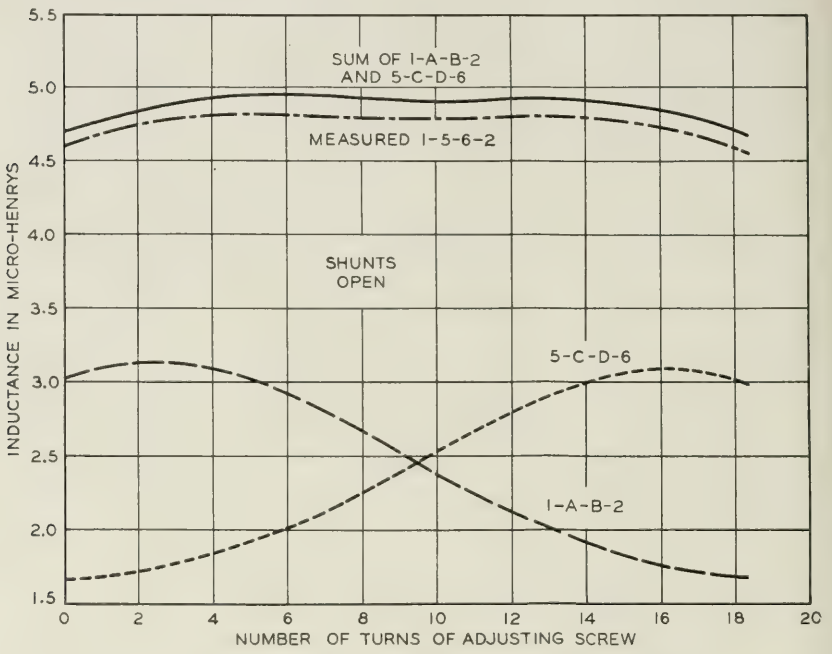


Fig. 10—Series inductance of non-shunted trial balancing coil.

Balancing Panels

In assembling the balancing coils on panels, the same number of coils should be traversed on each of two pairs before reaching the coil that balances these two pairs, in order that the phase shift up to this balancing coil on one pair will be essentially the same as that on the other pair. If these phase shifts differed materially, the coil setting for minimum crosstalk when one pair is the disturbing circuit might be quite different from the best setting when the other pair is the disturbing circuit. To obtain this equality objective a "criss-cross" arrangement, as shown schematically on Fig. 11, was devised, whereby the number of coils on one pair up to a particular balancing coil never differs by more than one from the number of coils on the other pair up to this same balancing coil.

For economic reasons it is undesirable to install a complete panel for the ultimate number of pairs, possibly 100 in some cases, but rather to install sections conforming more closely to the circuit growth. The placing at different times and properly connecting of sections obtained from the 100-pair criss-cross panel and at the same time maintaining service on operating circuits appeared rather formidable. This problem was solved by the use of two types of criss-cross panels; an

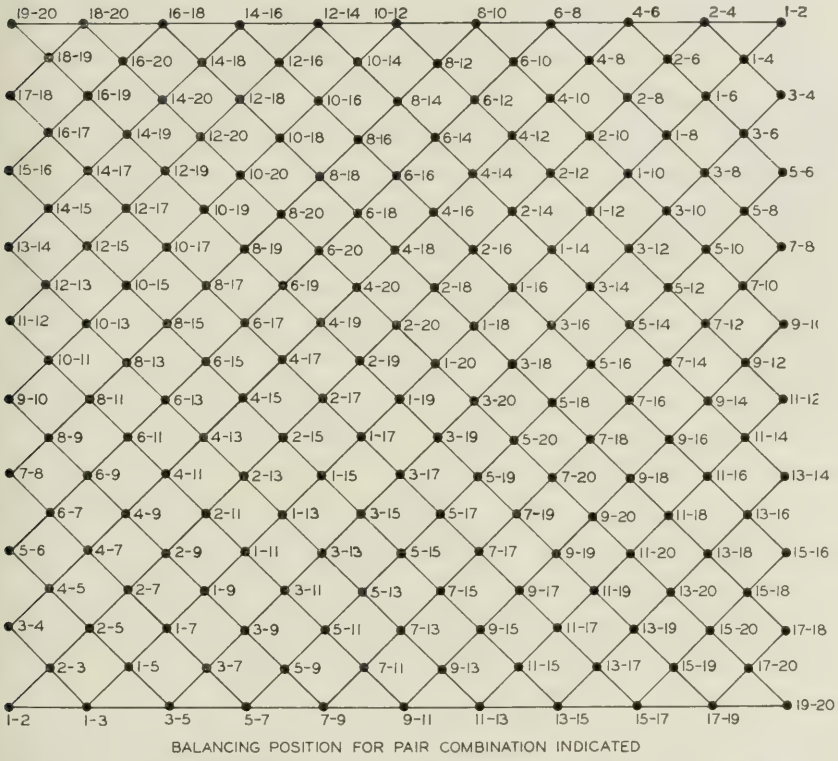


Fig. 11—Schematic of criss-cross wiring for 20-pair balancing panel, designed to maintain phase equality of coils.

intra-group panel for balancing within one group of carrier pairs and an inter-group panel for balancing pairs in one group against pairs in a second group of equal size. In the present design, an intra-group panel takes care of 20 pairs (190 combinations) and an inter-group panel of the 400 combinations between two 20-pair groups. To maintain phase equality through a number of panels, it is necessary to install them following a definite pattern. Figure 12 shows a suitable pattern for the 15 panels required for 100 pairs.

In the criss-cross scheme (Fig. 11) the side-to-side combinations, which are those marked 1/2, 3/4, 5/6, etc., appear twice, i.e., along the left and right edges of the panel. Advantage of this is taken by installing balancing coils at both locations. This is done because one side-to-side coil of about 1.3 microhenries may not be large enough in all cases in spite of the fact that the mean side-to-side crosstalk has been reduced 9 db by poling.

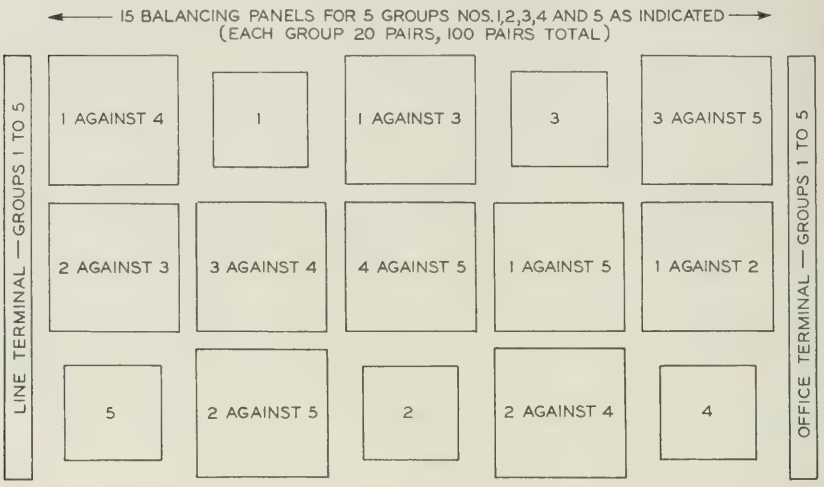


Fig. 12—Allocation of balancing panels designed to maintain phase equalization of coils at all stages. Panels with suitable cross-connections between them are installed in following order.

For first group—Install 1
Add second group—Add 2, and 1 against 2
Add third group—Add 3, 1 against 3, and 2 against 3
Etc.

Balancing Procedure

As stated above, the far-end crosstalk in a repeater section can not be balanced out completely over the frequency range with a single balancing unit. To determine the balanceable as distinct from the non-balanceable crosstalk, involves crosstalk measurements in phase and magnitude at a number of frequencies, using each pair of a two-pair combination as a disturbing circuit in turn. The balanceable crosstalk may then be separated from the non-balanceable crosstalk by computation. Balancing by this method would be impracticable because of the time required. As a practical scheme, it has been shown that balancing at a frequency of about 40 kc. will produce satisfactory results over the type K range even though part of the non-balanceable crosstalk may

be neutralized at this frequency. This is theoretically undesirable since the crosstalk reduction at other frequencies is impaired.

To prevent undue interference into operating carrier circuits when balancing, a frequency falling between the transmitted bands must be used. For this reason, the balancing coils are adjusted at a test frequency of 39.85 kc. and a measurement to check the suitability of the adjustment is made at 28.15 kc. Figure 13 shows the crosstalk vs. frequency before and after coil balancing by this method on three repeater sections.

Additional Crosstalk Remedial Measures

Although poling as well as balancing is done to reduce side-to-side crosstalk, this crosstalk is still considerably greater than the pair-to-pair crosstalk. For this reason, side-to-side crosstalk is diluted among

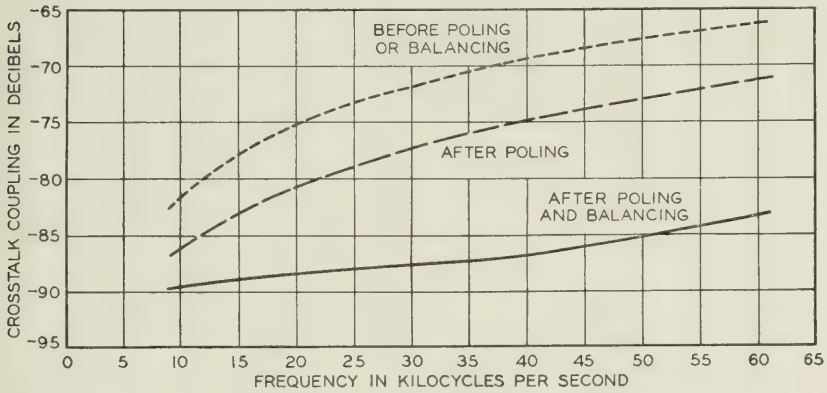


Fig. 13—R.M.S. far-end crosstalk per repeater section from measurements on 3 repeater sections.

the pair-to-pair combinations by a system of quad-splitting at repeater points.

The crosstalk after balancing (Fig. 13) is considerably higher at the upper end of the frequency band than at the lower end. Consequently, if circuits were set up to use the same channel throughout, the crosstalk in the upper-frequency channels would be materially greater than that in the lower-frequency channels. In order that all circuits may be equally satisfactory from the crosstalk standpoint, a system of special channel assignments in successive intervals, say 500 to 1000 miles, can be used. This will tend to equalize both the crosstalk and the noise on all circuits, thus permitting a somewhat cheaper design then if each channel had to meet the crosstalk and noise limits by itself.

Noise

Besides babble, many other sources of noise need to be considered in cable carrier design. Figure 14, which shows the approximate magnitude of several of these if no means are taken to suppress them, indicates the noise at the end of a single 17-mile repeater section when

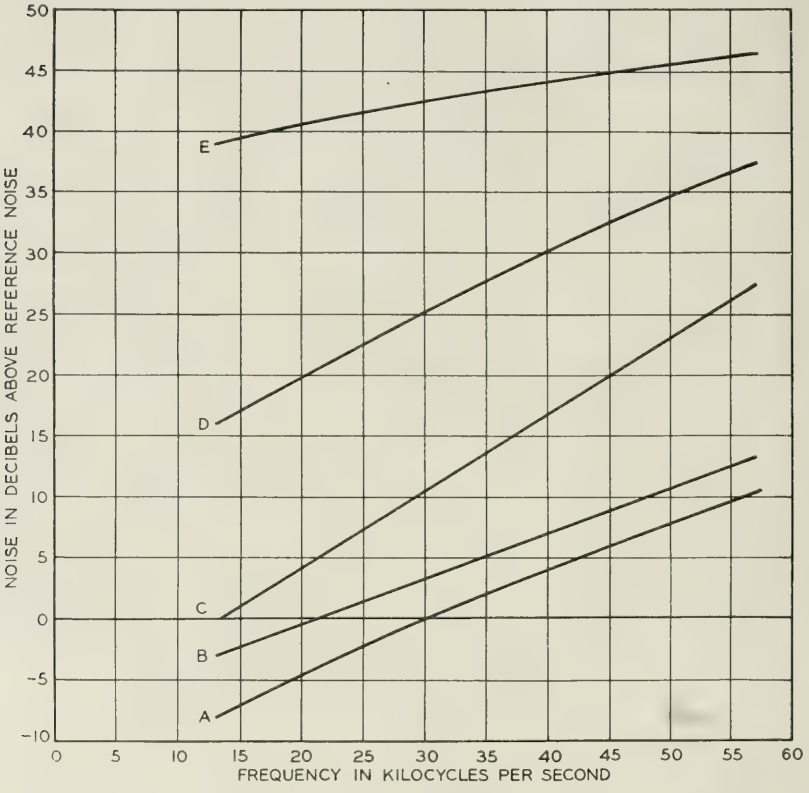


Fig. 14—Noise, prior to suppression measures, per repeater section at output of repeater whose gain equals line loss.

A—Noise from thermal agitation.
B—Thermal agitation plus tube noise.
C—Noise from voice frequency telephone repeater office.
D—Noise from telephone and telegraph repeater office.
E—Noise from heavy static on open-wire tap close to carrier repeater input.

amplified by a repeater whose gain equals the hot-weather line loss. Curve *A* shows the unavoidable lower limit of noise, that produced by thermal agitation of the electrons in the cable conductors and the repeater.[2] This amounts to about 2×10^{-17} watts per telephone

[2] J. B. Johnson, *Phys. Rev.*, **32**, 97 (1928); H. Nyquist, *Phys. Rev.*, **32**, 110 (1928).

channel per repeater section, at the repeater input. If there were no other noise sources, the repeater section length would necessarily be limited by this effect. Curve *B* shows the sum of thermal noise and noise due to the vacuum tubes in the repeaters, which is little in excess of thermal noise alone. The other three curves show noises of considerably higher magnitude which require suppression in order to arrive at an economical carrier system. Curve *E* shows the order of magnitude of noise on carrier circuits due to connecting open-wire pairs directly to non-carrier pairs in the outside cable near the carrier repeater input. The source of the noise is heavy atmospheric static of a magnitude experienced several times during the summer.

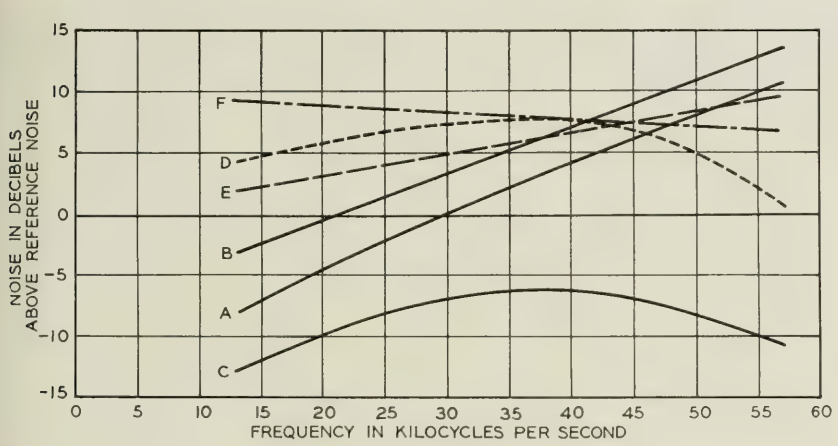


Fig. 15—Noise, subsequent to suppression measures, per repeater section at output of repeater whose gain equals line loss.

A to *E*—Same sources as in Fig. 14.

F—Noise from heavy static induced directly into outside cable.

The other curves show typical magnitudes of noise originating in the existing telegraph and voice frequency telephone plant; this is generated in existing repeater stations and transmitted by the non-carrier pairs to the outside cable where it is induced into the carrier pairs. Curve *D* represents the situation at a combined telephone and telegraph repeater station, and Curve *C*, the situation at a station where there are no telegraph repeaters.

Figure 15 indicates the results after suppression measures have been applied. As shown, at the top frequency, which controls the carrier repeater section length, these sources of noise have been reduced to be well below thermal plus tube noise. It is also shown that the noise due

to heavy atmospheric static induced directly into a carrier pair in the outside cable is below thermal plus tube noise at the top frequency.

There are additional types of noise, not shown, whose sources lie within the carrier system: e.g., modulation in amplifiers, inter-system cross-induction, battery noise. While control of such noise is an integral part of the fundamental carrier system design, it is not the purpose of this paper to cover this class of noise.

Conductors Tapping the Carrier Cable

Carrier noise may come from open-wire pairs which connect to conductors in the cable. Its sources may be static; corona on power lines; power line carrier or other carrier frequency voltages on power lines paralleling the open wire; induction from radio telegraph stations; or carrier frequency voltages arising in the office to which the open

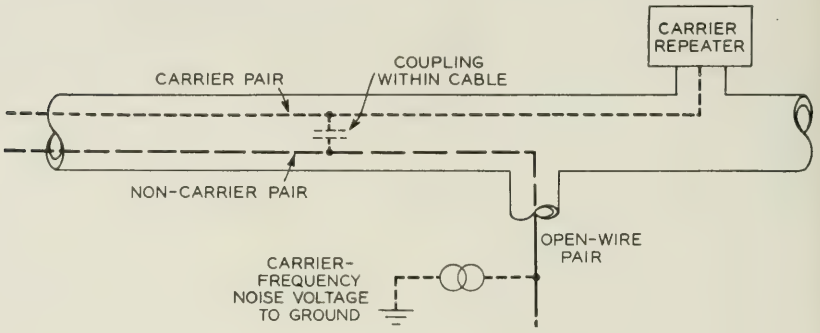


Fig. 16—Schematic of path followed by induction from open-wire taps.

wire is connected, such as voltages generated by d-c. telegraph or telephone signaling systems. The limited experience to date indicates that, in a long cable carrier system, the effect of heavy static will be larger than that of the other sources if telephone and power supply plants are coordinated so as to be satisfactory from voice frequency and low frequency standpoints. Branch cables connected to the carrier cable have a similar but generally smaller effect than that of open-wire taps.

Figure 16 illustrates the path followed by this induction. A voltage to ground impressed on the open-wire pairs passes by secondary induction over to the carrier pairs in the cable, and, on account of the unbalance to ground of these pairs, produces a metallic voltage on these pairs at the repeater input. The effect may be greatly reduced by interposing a filter at the junction of the open wire and the cable.

It is necessary to filter only the longitudinal circuit at an open-wire tap, because: (1) the voltage to ground on the open wire is larger than the metallic circuit voltage, and (2) the coupling between the longitudinal circuit and the disturbed carrier pair is greater than the coupling between metallic circuits.

Figure 17 is a schematic diagram of the longitudinal filter developed for a phantom group. It consists of two longitudinal retardation coils and a set of condensers connected between the line wires and the cable sheath. This filter has relatively high carrier frequency longitudinal impedance to minimize effects of impedance in the ground con-

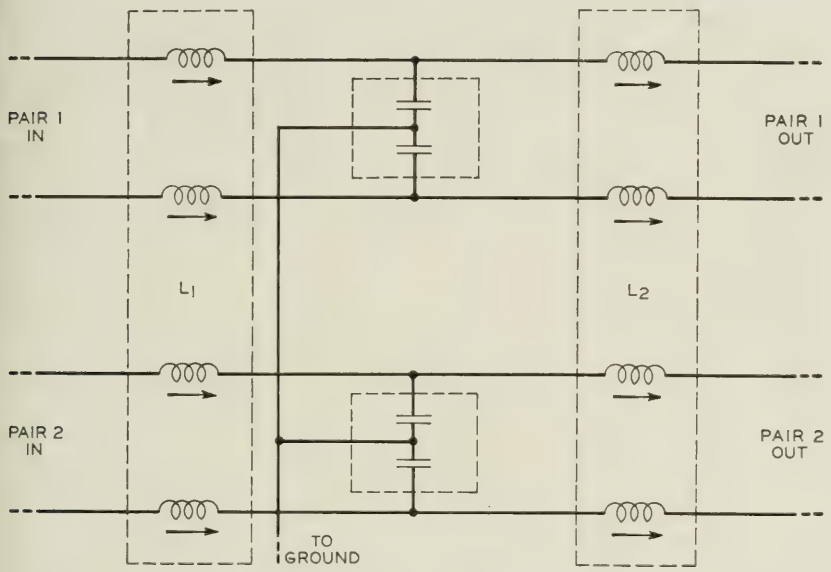


Fig. 17—Schematic of longitudinal filter.

nection. The major portion of the carrier frequency impedance of the coils is obtained by designing them to have high core loss at these frequencies. The filter has little effect on voice frequency transmission, precaution having been taken to hold the transmission loss, crosstalk and unbalance to ground to low values.

Noise Arising in Existing Repeater Offices

The noise caused by carrier frequency voltages generated in existing repeater offices is due to d-c. telegraph, telephone speech and signaling voltages, power supply, etc. Figure 1 shows the path by which they reach the carrier plant and the means used to suppress them. In this

figure, N represents a source of carrier frequency voltage in a repeater office, connected to a voice frequency pair which transmits this voltage into the outside cable where it is induced on the carrier pairs. These voltages are reduced by inserting suppression coils in the longitudinal voice frequency paths at the junction between the office and the outside cable connected to carrier inputs.

The design of coils giving the requisite carrier frequency suppression without appreciably affecting voice frequency transmission on the circuits in which they are connected was difficult. One coil is used for each phantom group. Each coil has sixteen windings, four for each line wire. These windings are so paired and disposed about the core

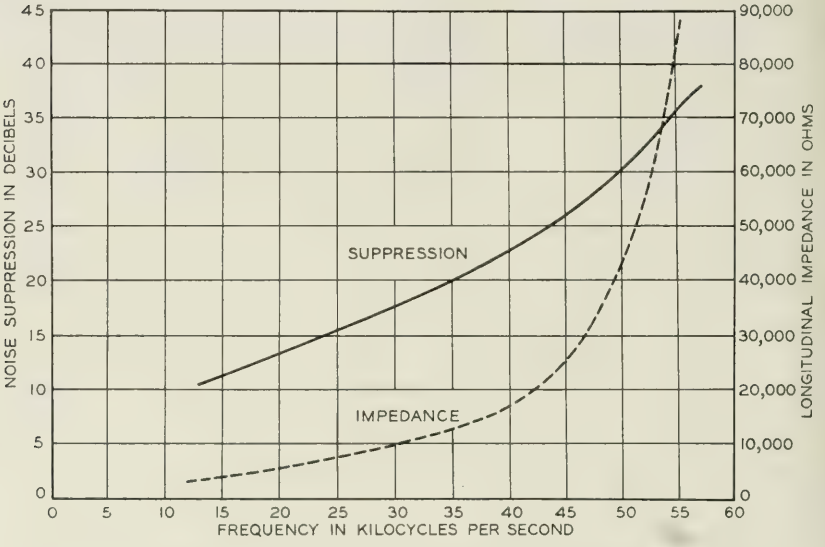


Fig. 18—Longitudinal impedance and suppression of noise suppression coils.

as to make possible very small side-to-side and phantom-to-side cross-talk between line windings. They also permit obtaining very small leakage flux in both the sides and the phantoms; hence the coils introduce very small transmission loss in their voice frequency circuits. The leakage impedance of the coils plus the impedance of the cable stub used to connect them into the circuit is held down so that the effect on repeater singing and echoes in the voice circuits is very small. The coils are so wound that their longitudinal inductance is in anti-resonance with their distributed longitudinal capacitance at approximately the top cable carrier frequency, resulting in a large increase in their suppression in this critical frequency range. The longitudinal

impedance of one of these coils, and the approximate suppression which a set of them provides, are shown in Fig. 18.

In addition, the carrier circuits are carefully separated, electrically and physically, from existing voice frequency circuits in common repeater stations. To this end the carrier pairs in the outside cable are brought out on the line side of the noise suppression coils into a separate cable connected directly to a sealed terminal. From this terminal they are carried in shielded wire to the units in the carrier office and then to a similar sealed terminal leading to the outside cable in the opposite direction. Filters for filament and plate battery supply are included in the carrier amplifiers and additional filament battery supply filters are provided at the carrier fuse panels.

A New Single Channel Carrier Telephone System *

By H. J. FISHER, M. L. ALMQUIST and R. H. MILLS

The single channel carrier telephone system described in the following paper is designated the Type H. It is characterized by several new features, making it applicable not only to the needs of telephone companies but also to those of railroads, power systems and oil companies. It replaces the Type D single channel carrier system, more than 500 of which are now in operation in the Bell System, and, in addition because of its lower cost is applicable to shorter distances. It therefore marks another step in extending the use of carrier. Reduction in size and provision for operating on a-c supply simplify its installation, and its portability makes it well suited to provide emergency circuits.

ON open-wire lines where the growth is not rapid, there is frequently need for adding telephone circuits one at a time. When the Type D single-channel carrier telephone system was developed a few years ago it became possible to meet this need without stringing additional wires.[1] More than 500 of these systems have been placed in service in the Bell System plant. A new single-channel carrier telephone system, known as the Type H, has recently been developed and is now being applied. This new system offers improved performance, and also, because of its lower cost, is applicable to providing service over shorter distances than were economical with the earlier system.

The Type H system, which is characterized by a number of new features and special developments, is applicable not only to the needs of telephone companies but also to those of railroads, power systems, and oil companies.[2] In the first place it is designed to operate either on alternating current or on direct-current plate and filament supply. A repeater is available to extend the range of operation. Through the use of specially designed but simple filters the system can be employed on circuits which are equipped with bridged telephone

* Presented at Winter Convention of A. I. E. E., New York, N. Y., January 24–28, 1938. Published in *Electrical Engineering*, January 1938.

[1] "Carrier Telephone System for Short Toll Circuits," H. S. Black, M. L. Almquist and L. M. Ilgenfritz, *A. I. E. E. Transactions*, Vol. 48, January 1929, pp. 117–140.

[2] "Carrier Telephone Systems—Application to Railroad Circuits," H. A. Affel, *Proceedings of the Association of American Railroads, Telephone and Telegraph Section*, October 1936, pp. 654–672.

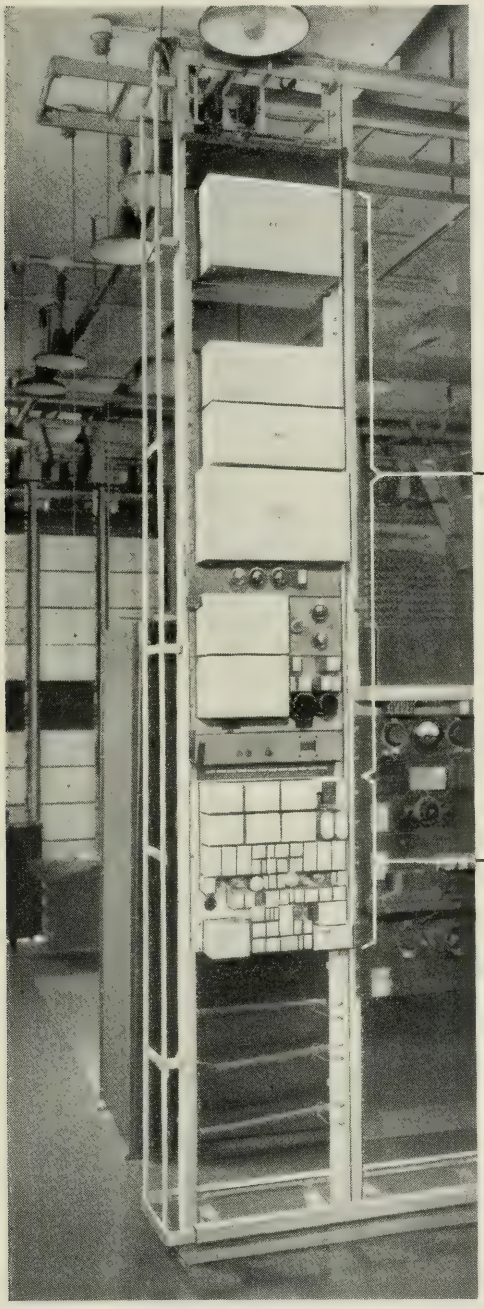


Fig. 1—Installation of Type H carrier telephone system at Charlotte, North Carolina.

stations at intermediate points as is frequently the case in railroad operation.

A unique feature is the use of opposite sidebands of the same carrier frequency for opposite directions of transmission. The upper sideband is used in one direction and the lower sideband in the other, the carrier being suppressed. For the modulators and demodulators copper oxide "varistors" are employed in place of vacuum tubes. The amplifiers are single stage, employ pentode tubes and are stabilized in performance by feedback. The filters have been simplified in construction by the use of coils with a new type of core material and by improved designs of paper condensers.

The size of the new terminal has been so reduced that it occupies less than 40 per cent of the space required for a Type D terminal, as indicated in Fig. 1. The equipment may be mounted on racks as is customary in telephone offices, or a complete terminal or repeater may be mounted in a small cabinet.

Single-channel carrier systems have been used in the Bell System principally for short open-wire toll circuits. Thus, the Type D systems are for the most part between 50 and 200 miles in length. The Type H system, since it includes a repeater, can be used for greater distances, and due to its lower cost is economical for shorter distances.

General Description of System

Basic System

The basic system consists of two terminals one of which is referred to as an "east" terminal and the other as a "west" terminal, as indicated in Fig. 2. The two terminals differ only in minor respects, the differences being due to the fact that at one terminal the upper sideband is transmitted and the lower sideband is received, while at the other terminal the reverse takes place. In order to simplify coordination between various types of carrier systems operating on the same pole line, the frequencies between 7400 cycles and 10,150 cycles are transmitted in the east to west (or north to south) direction, and the frequencies between 4150 cycles and 6900 cycles in the west to east (or south to north) direction. The frequency allocation of the Type H system and those for the Type D and the three-channel Type CS system are shown in Fig. 3. All three types may be operated on the same pole line.

The circuit arrangement is given in greater detail in Fig. 4, which shows a schematic diagram of one terminal, with the exception of the power supply circuit. Each terminal is made up of a transmitting

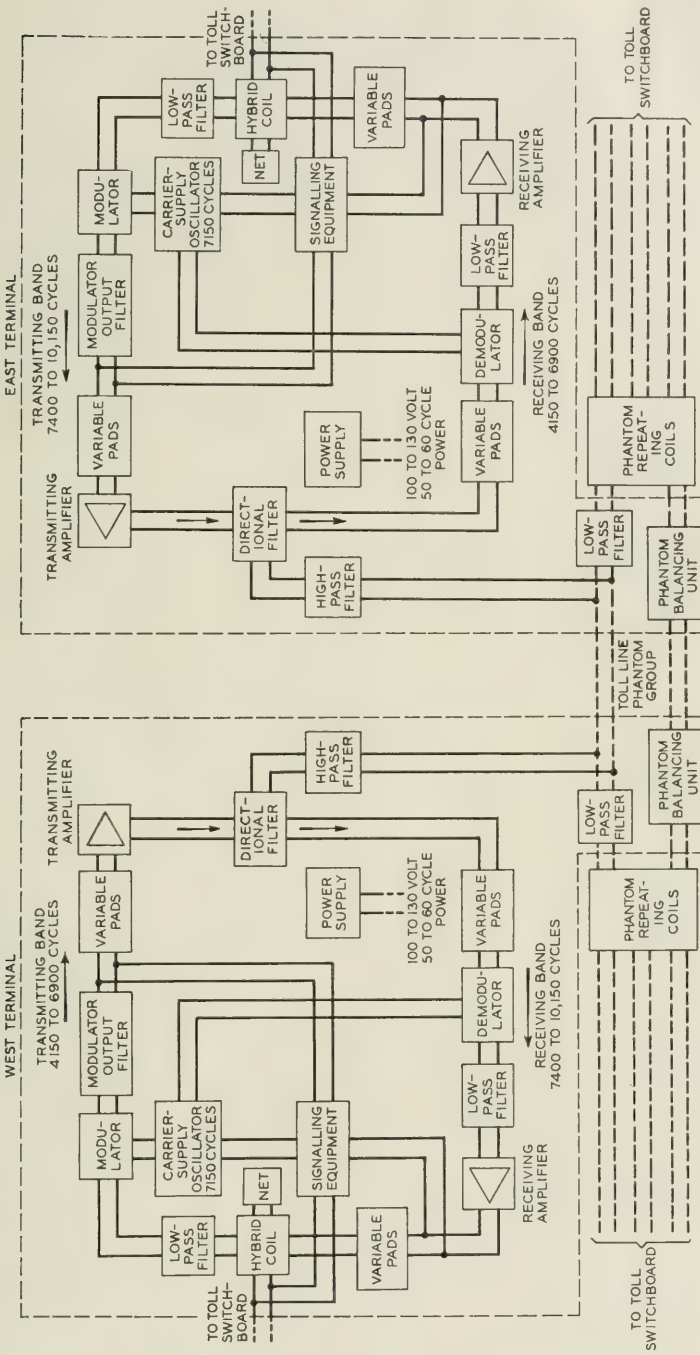


Fig. 2—General schematic of Type H system applied to side circuit of phantom group.

branch which includes a modulator, a receiving branch which includes a demodulator, and a hybrid coil to combine the two branches into a two-wire voice-frequency circuit. The carrier is generated by a vacuum tube oscillator which supplies both the modulator and demodulator. The output of the modulator, which is of the copper oxide type, consists principally of the two sidebands. The desired sideband is selected by the modulator output filter and an amplifier raises the level to that desired for transmission over the line. The demodulator is also of the copper oxide type; its output consists principally of the two sidebands, one of which is a reproduction of the original voice-frequency input. This is selected by means of a low-pass filter and applied to a voice-frequency amplifier which provides the necessary receiving gain. Adjustable pads serve as a means for adjusting the transmitting and receiving gains. The characteristics and functions of the various filters are described later.

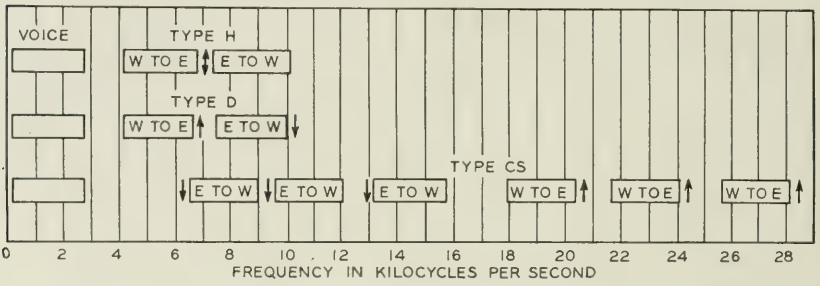


Fig. 3—Frequency allocation.

For signaling over the carrier circuit, a 1000-cycle signaling system is employed. The method is similar to that used on other types of circuits, but includes a number of simplifications. The number of tubes has been reduced, and the vibrating relays used in the older type circuits have been replaced by copper oxide rectifiers and simple d-c. relays requiring little maintenance. When 20-cycle ringing current is received from the switchboard, the frequency of the carrier supply oscillator is shifted by 1000 cycles and its output is interrupted at a 20-cycle rate, and applied to the input of the transmitting amplifier. At the receiving end this appears at the input of the signal receiving circuit as a 1000-cycle current interrupted at a 20-cycle rate. It is then demodulated in a copper oxide rectifier and the resulting 20-cycle current is amplified and applied to a second rectifier the output of which is connected to a d-c. relay. Operation of this relay causes 20-cycle current from a local source to be sent toward the switchboard.

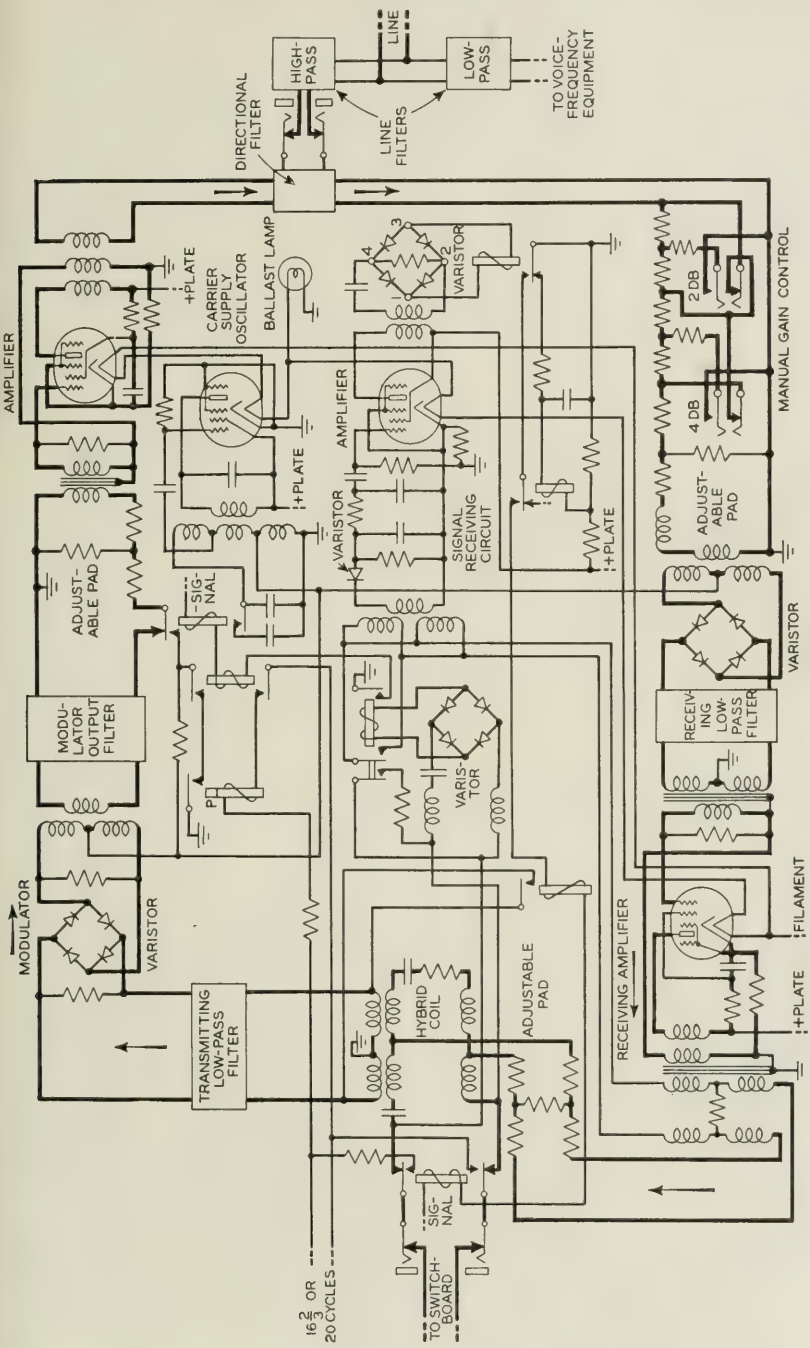


Fig. 4—Schematic of Type H terminal.

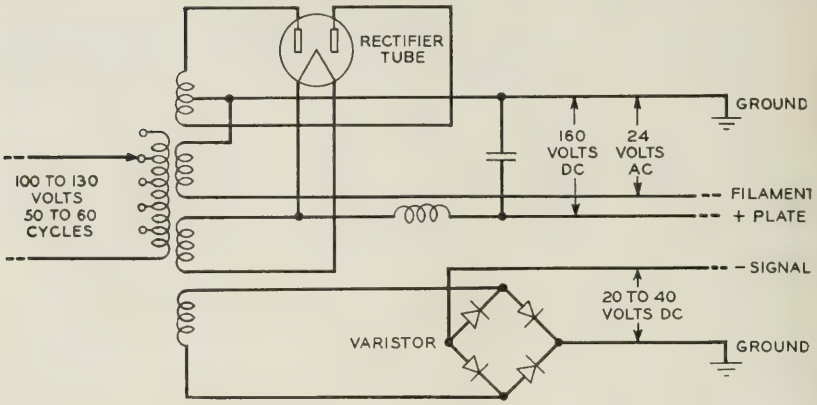


Fig. 5—Schematic of a-c. power supply for terminal.

Freedom from false operation on speech is obtained by tuning the receiving circuit to the rate of interruption and by providing a slow operating relay such that the signal must persist for several tenths of a second before 20-cycle current is passed to the switchboard.

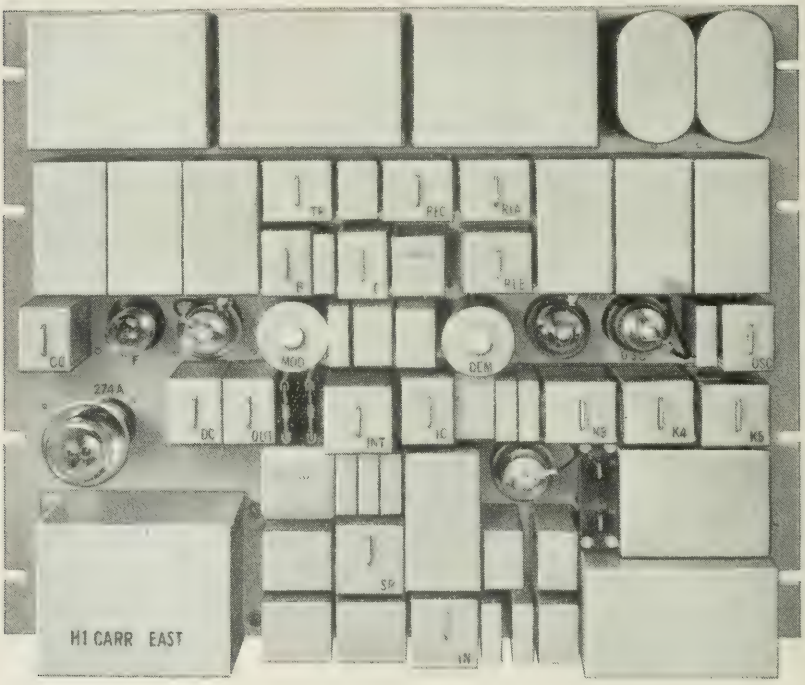


Fig. 6—Terminal panel—front view.

The power supply circuit, which is shown in Fig. 5, operates from an alternating-current source of 100 to 130 volts, 50 to 60 cycles, and requires about 50 watts. It supplies alternating current for the filament supply, 160 volts direct current for plate supply and 24 volts direct current for relay operation. Provision is also made for direct operation from 24-volt and 130-volt central office batteries.

The equipment for a terminal is mounted on two panels. One of these, shown in Fig. 6, is $15\frac{1}{4}$ inches by 19 inches in size and contains all the apparatus except the line filter circuit. The second panel, shown in Fig. 7, is $3\frac{1}{2}$ inches by 19 inches and contains the line filters, and other equipment which is required for balancing purposes.

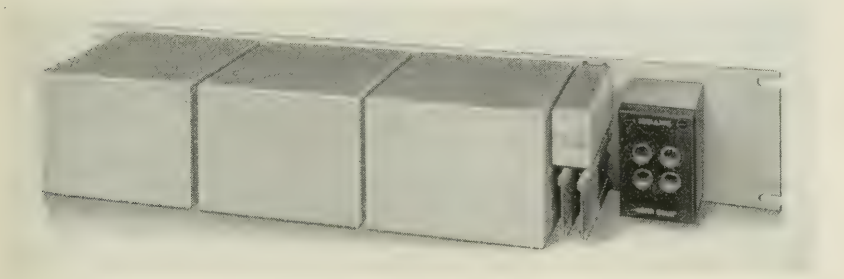


Fig. 7—Line filter and balancing panel—front view.

Repeater

A schematic of the repeater is shown in Fig. 8. The amplifiers are the same in design as the transmitting amplifier at the terminals. Two sets of line filters are required and are identical with those used at the terminals. The a-c. power supply circuit is substantially the same as the one used at the terminals except that the 24-volt supply for relay operation is omitted. The power required for operation is about 35 watts.

The complete repeater consists of three panels—the repeater panel, which is $10\frac{1}{2}$ inches by 19 inches, and two line filter and balancing panels. The repeater panel is shown in Fig. 9.

Transmission Performance

Line Considerations

In outlining the transmission performance of the system, it is necessary to consider the characteristics of the lines as well as those of the equipment. At carrier frequencies the line losses and the variations in these losses with weather are considerably greater than at voice frequencies. This is illustrated in Fig. 10, which shows the

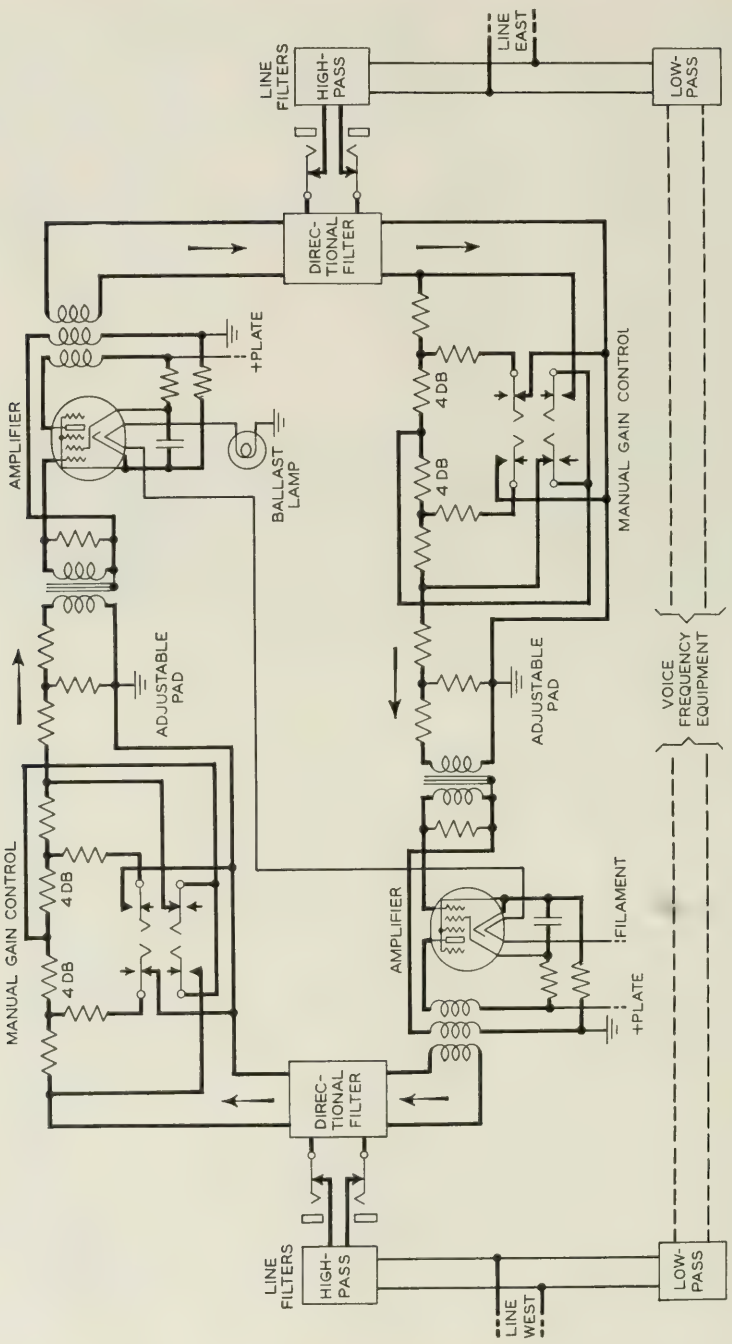


Fig. 8—Schematic of repeater.

attenuation-frequency characteristics for four commonly employed gauges of open-wire line under dry and wet weather conditions, respectively, at a temperature of 68° F. These curves are for 12-inch spaced copper wires equipped with the type of insulators ordinarily

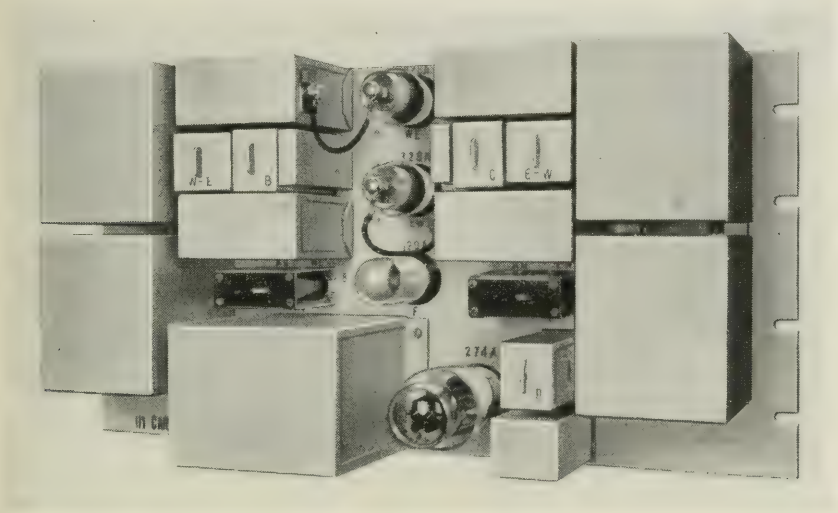


Fig. 9—Repeater panel—front view.

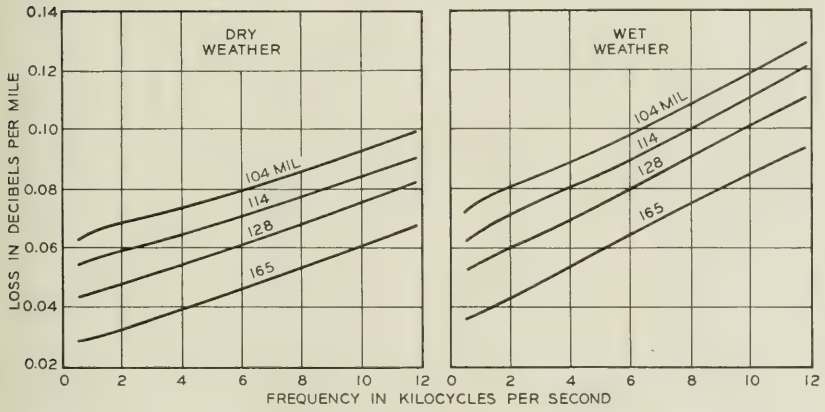


Fig. 10—Attenuation of open-wire side circuits—12-inch spaced copper wire with double petticoat glass insulators.

employed for toll circuits. Changes in temperature also affect the attenuation, the loss increasing as the temperature rises. The variations due to changes in temperature are, however, smaller than those due to changes from dry to wet weather and are more likely to occur

gradually. It is apparent that the variations in attenuation with weather increase with the length of the system, and that more frequent adjustments will be required on the longer systems to maintain the overall circuit net loss within given limits.

The carrier frequency attenuation of cable circuits is much greater per unit length than that for open-wire lines. Hence, the losses introduced by comparatively short sections of cable, either at entrances to offices or at intermediate points, are of considerable importance. In addition to the attenuation of the cable itself, there are large reflection losses at the junction of the cable and the open wire, due to the difference between the characteristic impedance of the open-wire and that of the non-loaded cable. The carrier terminal and repeater have been designed to have the same impedance as the average of the open-wire line facilities, so the same considerations apply at the junction of the cable and the carrier equipment. As an example of the magnitude of these effects, the insertion loss at 8000 cycles of two miles of non-loaded 16-gauge cable is approximately 6 db, or about the same as that of 60 miles of 104-mil open-wire.

By applying carrier loading that has been developed for this purpose, the attenuation loss of such a cable can be reduced to about 1 db at 8000 cycles. In addition, the reflections at the terminals of the cable will be reduced to satisfactory low values by virtue of the impedance matching properties of the loading. This method of treatment has the important advantage that it improves the transmission characteristics in both the voice and carrier ranges.

In cases where substantial transmission margins exist, it is sometimes practicable to use impedance matching transformer networks at the cable terminals, as a substitute for carrier loading, with economies that are proportional to the length of the cables. At the present stage of development, this so-called transformer treatment is much less satisfactory in the voice-frequency range than in the carrier-frequency range. Certain inherent limitations in simple transformer treatment result from the fact that the ratio of the (non-loaded) cable impedance to the open-wire impedance varies widely over the frequency band to be transmitted, and the transformer impedance ratio that is optimum at carrier frequencies is distinctly disadvantageous in the voice-frequency range. The choice of optimum transformer ratio for the complete transmission band may thus involve different compromises for different sets of conditions and service requirements.

Where several carrier systems are to be placed on a pole line, crosstalk between systems becomes an important consideration. Where

only a few single-channel systems are involved, it is sometimes possible, by separating the systems widely on the pole line, to operate with only the regular voice-frequency transpositions, but, in general, additional transpositions are required. A comparatively inexpensive transposition system for this purpose was designed at the time the Type D system was developed. It permits operation of Type H systems on all pairs of a four-crossarm line with the exception of the pole pairs, and Type C three-channel systems on the top crossarm. In addition to transposing it is important that reflections at junctions between open wire and cable be reduced as described above, in order that near-end crosstalk will not through reflection appear as crosstalk at the distant terminal of the system.

Range of Operation

The terminals and repeaters have sufficient load carrying capacity so that they may be operated at an output level 16 db above that at the transmitting toll switchboard. About 19 db transmitting gain and 14 db receiving gain are available at each terminal, of which a total of 20 db may be used at the east terminal and 22 db at the west terminal. The lower permissible loop gain (sum of transmitting and receiving gains) at the east terminal is not controlling, since the line loss is greater for the frequencies used in the east to west direction than for those used in the west to east direction. Thus, the terminals are capable of providing a 9 db circuit over a line the attenuation of which does not exceed 31 db at 8150 cycles and 29 db at 6150 cycles. These figures correspond roughly to the wet weather attenuation of about 280 miles of 104-mil open wire where no intermediate cable or equipment is involved. The presence of even a small amount of cable will considerably increase the attenuation so that in most cases the distance which can be spanned is not greater than 150 to 200 miles, and may be even less.

Where greater distances are to be covered, an intermediate repeater may be added. The repeater has a useful gain of about 23 db in each direction, with some flexibility as to allocation of gains between the two directions of transmission. More than one intermediate repeater can, of course, be employed, although as the system is lengthened maintenance effort will be increased, as more frequent adjustments of the overall net loss will be required to compensate for the variations in line attenuation with weather. No provision is made for a pilot channel such as is generally provided on the long multi-channel systems, and adjustments of overall net loss must be made manually. Also, no provision has been made for equalizing the variation in line

attenuation with frequency. These factors are not important on the shorter circuits for which the system has primarily been designed.

Overall Transmission Characteristics

The circuit provided by a Type H system without a repeater has a band width of about 2750 cycles, extending from about 250 to 3000 cycles. This is somewhat wider than that for the Type D system. The introduction of repeaters will tend to narrow the band somewhat. Representative frequency characteristics are shown in Fig. 11. One

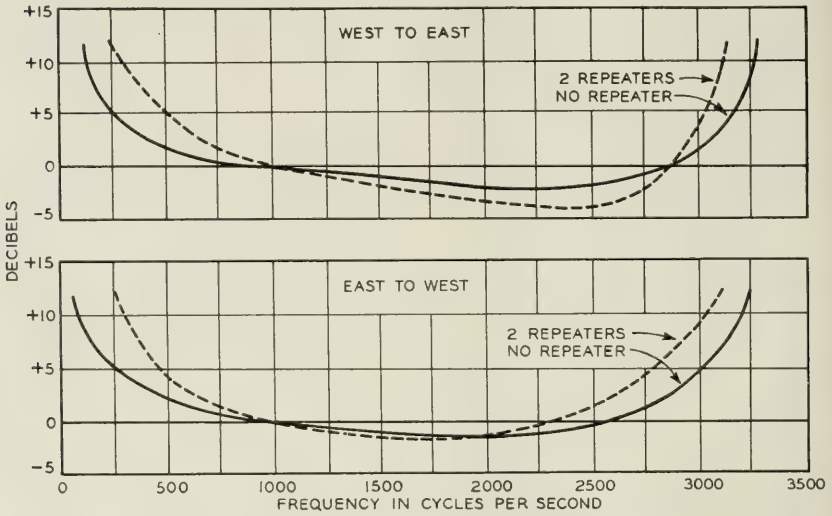


Fig. 11—Representative overall circuit net loss characteristic of system.

set of characteristics is for a typical circuit without a repeater, and the other for a circuit including two repeaters. It is assumed that the line conditions are such that no large reflection effects are present.

The band width is limited principally by the characteristics of the band filters. The small differences in the characteristics for the two directions of transmission are due partly to differences in the filters and partly to the fact that the attenuation of the line increases with frequency and is greater at the 3000-cycle point in the east-to-west channel than for the 3000-cycle point of the west-to-east channel. As the circuit is increased in length this difference tends to increase.

Variations in overall circuit net loss are due largely to variations in the loss of the high-frequency line. For a circuit 200 miles long these may amount to $\pm$ 3 db. The key controlled pads which are included at each terminal and repeater are provided for making adjustments to

compensate for these variations. Variations due to the equipment are small in comparison with the line variations. The transmitting gain at a terminal may vary ± 0.5 db and the receiving gain ± 0.3 db for variations of ± 10 volts in the a-c. supply. With a more stable a-c. supply or when operated from regulated plate and filament batteries such as are employed in the larger telephone offices, these variations will be less than half the figures given above. With suitable maintenance it should be possible to maintain the overall circuit net loss within ± 2 db of its normal value.

A representative load characteristic, as measured with 1000-cycle current for a system without a repeater, is shown in Fig. 12. On a

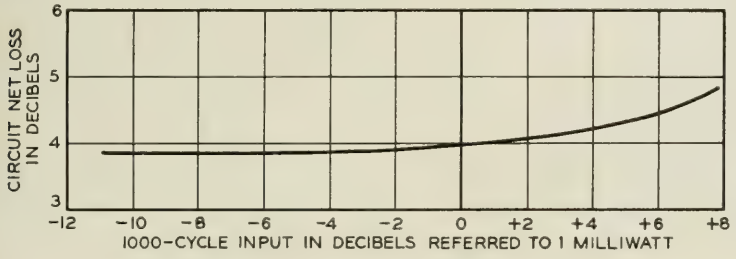


Fig. 12—Representative system load characteristic.

repeatered system some additional limiting of high inputs may occur. However, even on repeatered systems, there should be no noticeable distortion for input volumes such as are obtained directly from a switchboard.

Reactions on Voice Frequency Circuit

In superimposing a carrier system on a voice-frequency circuit, line filters are added to provide separate paths for the voice and carrier circuits. The introduction of a filter in one side of a phantom group requires the addition of a network on the other side to maintain the balance of the phantom circuit from a noise and crosstalk standpoint. It is also necessary for return loss reasons to balance these units in the network circuits of voice-frequency repeaters that may be located at the same point as a carrier terminal or repeater. The networks required to take care of these conditions are included with the carrier system.

In some cases it is desired to apply the Type H system to circuits equipped with bridged telephone stations at intermediate points. Such arrangements are common on railroad communication systems, and occur to a small extent in the Bell System. In such cases, ex-

cessive interference into the carrier system due to talking at the way station, and into the way stations due to talking on the carrier system, is likely to occur unless suitable filters are provided at each way station. A simple filter for this purpose has been developed for use with the 501 type subscribers set, and work is proceeding on a similar filter for use with other types of subscriber sets.

The line filters and the filters for use at way telephone stations each introduce a loss of about 0.15 db to the through voice-frequency transmission. Where the voice-frequency circuit is equipped with repeaters and return loss conditions permit, these additional losses may be taken care of by readjusting the voice-frequency repeater gains. In extreme cases, particularly where a considerable number of filters are to be added, it may be necessary to resort to other means of improving the transmission on the voice-frequency circuit, such as loading of incidental cables or the addition of a voice-frequency repeater.

On circuits equipped with way stations, selective signaling by means of selectors is sometimes used. Such signaling systems are generally arranged to apply an "answer back" tone to the line when a station has been called to indicate to the calling party that the selector has operated. This tone contains a considerable amount of high-frequency currents so that it is necessary to modify the selector circuit to filter out the high frequencies. The modification is a simple one and makes the answer back tone inaudible on the carrier circuit.

Design Features

In the development of the Type H system advantage was taken of many new devices which have been perfected in recent years, adapting them to the particular conditions of this application. A discussion of the more interesting features relating to the design of the various parts of the system is given below.

Modulators

The modulator and demodulator used in the Type H system are of the double-balanced copper-oxide type. Each modulator or demodulator consists of an input transformer, an output transformer, a copper-oxide "varistor" and a carrier supply. Although the modulators are bilateral, in the present application they are used in one direction only. The varistor consists of 48 copper-oxide discs assembled on a single bolt and connected as shown in Fig. 13.

The principal advantage of this type of modulator or demodulator is that in the ideal case (and to a lesser degree in the practical case)

each modulation product appears only in one of the four branches of the circuit. For example in the case of the modulator, if a voltage of frequency "V" is applied to the input and a voltage of frequency "C" is applied by the carrier supply circuit, resulting products of modulation will appear in the ideal case as shown in Fig. 13. It is obvious that the only unwanted products in the output which cannot be suppressed by filters or balance are those which are of the frequency $(C \pm AV)$ which for some values of A and V fall in the frequency range of the desired sideband $(C + V)$ or $(C - V)$.

These components, however, are normally more than 50 db weaker than the sideband and are not noticeable. Of course, the term AV represents not only odd harmonics of V but odd order intermodulation

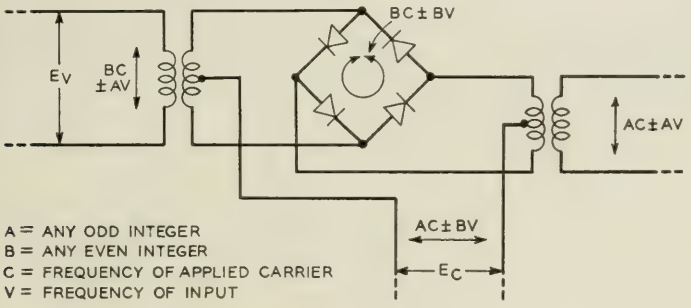


Fig. 13—Simplified schematic of copper oxide modulator.

products such as $(2V_1 \pm V_2)$. The relative amplitudes of the $(C \pm AV)$ terms increase with load in a similar manner to that for the distortion products of an amplifier as the overload point is approached, and the effect on articulation is the same.

For the actual case the modulator balance is not perfect and all products and the original frequencies do appear in all branches of the circuit including the output. However, the balance in most cases is greater than 30 db and the filter requirements are helped to that extent over some portions of the frequency range. This is particularly helpful in connection with suppressing the carrier from the output and input since it lies only about 200 cycles from the pass band and it would be costly to obtain all of the suppression required by means of filters.

With a single disc in each arm, taking the factory run of discs and making no attempt to select units, the balance for many assemblies would be less than 15 db. By selecting units, this balance could be improved to any desired amount. In the present design, however, to

save the cost of selection, twelve discs were used per arm to obtain the better balance resulting from the averaging of the characteristics of a large number of discs. There is some sacrifice in efficiency due to using the large number of discs but in this application it was of minor importance.

The averaging obtained by using twelve discs in each arm is helpful in several other respects. First, the normal impedance, transmission and balance do not vary greatly from unit to unit. Secondly, although each disc has a negative coefficient of resistance *vs.* temperature and there is a variation in the coefficient among discs, the average coefficient of twelve discs chosen at random will be very nearly equal to the average of any other twelve discs chosen at random and the balance between arms will, therefore, remain practically constant with temperature, even though the impedance and efficiency vary. A similar advantage is obtained in the case of aging and a good balance is obtained throughout the life of the equipment.

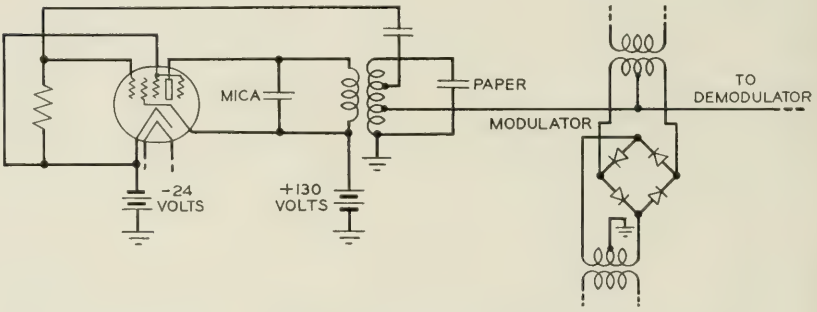


Fig. 14—Schematic diagram of oscillator circuit.

Oscillator

As mentioned previously the same carrier (suppressed) frequency is used for transmission in both directions, thus requiring a single oscillator instead of two as in the Type D system. The principal requirement for an oscillator for this use is that its frequency remain stable under the variations in power supply voltage and temperature which will occur. The new design, which is shown schematically in Fig. 14, provides a degree of stability such that no operating adjustments will be required due to these factors. Relatively high stability with changes in temperature is obtained by balancing the positive capacitance-temperature coefficient of the copper-oxide load and the mica tuning condenser against the negative coefficient of the paper tuning condenser.

The stability of frequency with plate voltage variations is about $5/10^6$ parts per volt. This is adequate and was obtained without the use of an expensive tuning inductance. The coil used, which also serves as output and feedback transformer, has a ratio of reactance to resistance of about 20 and is an air-core solenoid potted in a copper can.

Amplifiers

Both the receiving and transmitting amplifiers employ a single pentode with about 9 db feedback. For this amount of feedback, the variations of gain and impedance due to power supply variations are reduced to at least one-third of the amount of the variation obtained without feedback, and the load-carrying capacity is increased about 1 db.

The two amplifiers differ in that the frequency range transmitted is different and in that the output transformer of the receiving amplifier also acts as an inequality ratio hybrid coil to separate the receiving

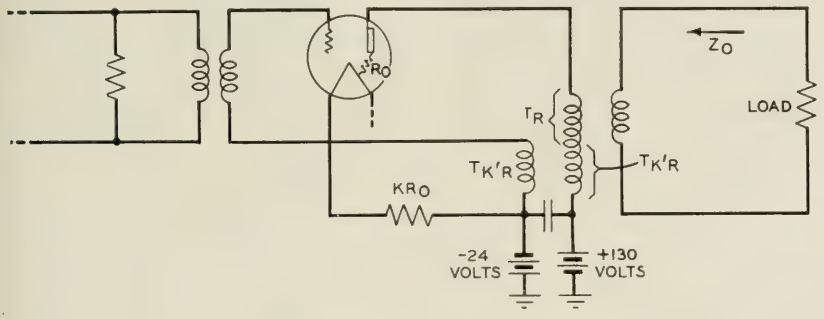


Fig. 15—Simplified schematic of amplifier.

signaling circuit from the two-wire voice circuit. The two circuits are shown in Fig. 4. In each case, the feedback is accomplished by means of a bridge circuit in the output and a series connection in the input. This can be more readily seen from Fig. 15, which is a simplified circuit representing both amplifiers. There is a considerable saving in circuit elements as compared to the familiar resistance bridge feedback connection. The output power loss due to shunt arms of the resistance bridge is eliminated. Furthermore, the impedance of the feedback circuit is relatively low, and consequently some wiring difficulties were avoided. In this application, the bridge is unbalanced, and the impedance Z_0 is a function of KR_0. As a result, it was convenient to adjust Z_0 to the optimum value by choosing the proper value of KR_0.

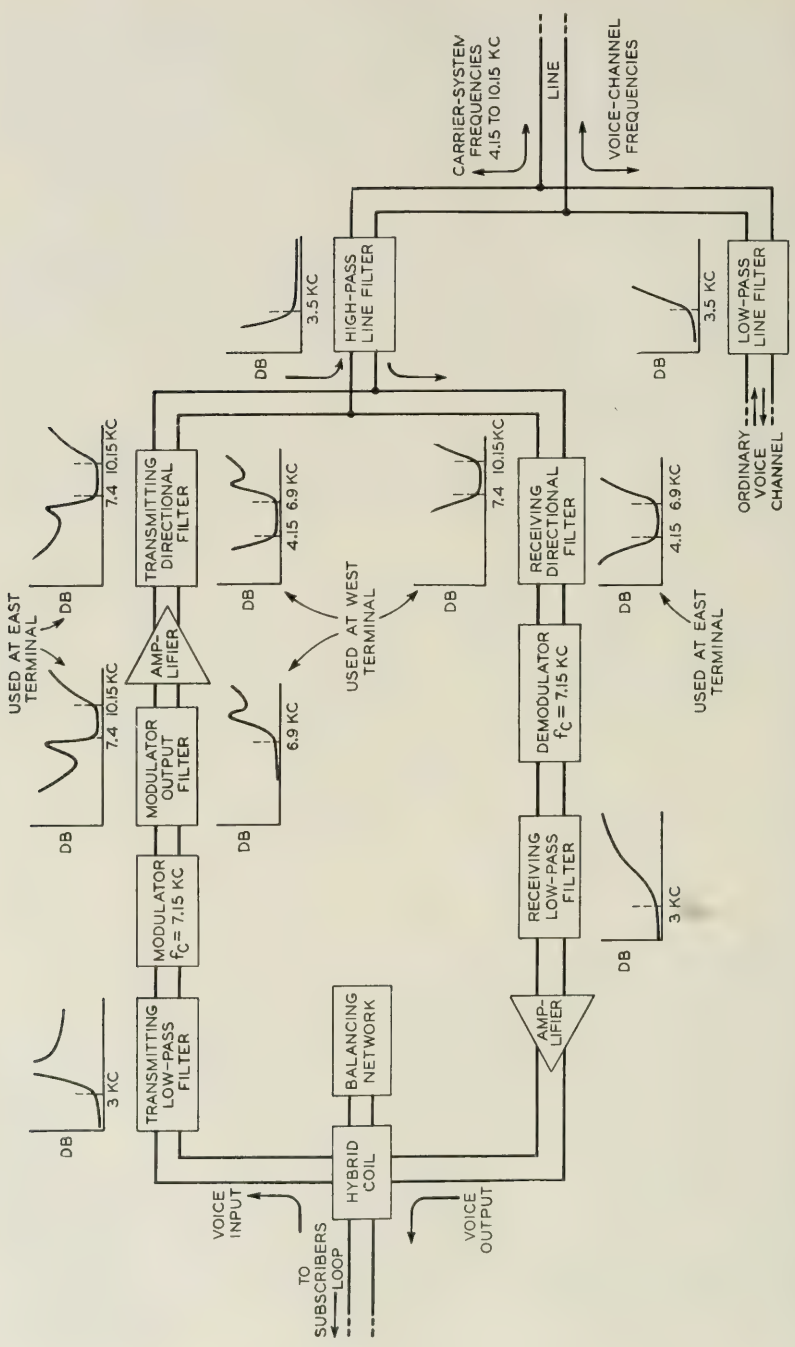


Fig. 16—Attenuation characteristics of filters at terminal.

Filters

Filters constitute an important part of the Type H carrier system. They represent about 30 per cent of the cost of the terminal and occupy about 25 per cent of the total space.

The various filters required at a terminal are indicated in Fig. 16. The transmission characteristic of each filter is given in miniature above or below the block representing the filter.

The high-pass and low-pass line filters separate the ordinary voice telephone channel and the added carrier telephone channel made available by this system. The low-pass line filter passes voice frequencies and suppresses all other frequencies. The high-pass line filter passes the carrier frequencies and suppresses the voice frequencies. Each filter offers a high impedance to the frequencies passed by the other, and bridges off only a very small part of the energy of these frequencies.

The remaining filters are associated with the carrier terminal proper, where they serve to separate the transmitting and receiving paths and suppress unwanted frequencies. The voice frequencies pass through the hybrid coil and the transmitting low-pass filter to the modulator. This filter limits the path between the hybrid coil and the modulator to voice frequencies only. Modulation of the voice with the carrier frequency of 7.15 kc. produces two sidebands extending from 4.15 to 6.90 kc. and from 7.40 to 10.15 kc. At an east terminal, the upper sideband is transmitted, and the modulator output filter passes this sideband and suppresses the lower sideband, together with other unwanted modulation products. In this manner it limits the load on the amplifier to the desired sideband. The transmitting directional filter offers further suppression to frequencies lying outside this band. The receiving directional filter will not pass this band but has a high impedance to these frequencies. The high-pass line filter passes all frequencies above roughly 3.5 kc. and, therefore, this band passes through it readily and out onto the line for transmission to the distant terminal. Transmission from a west terminal is identical in principle but here the lower sideband is passed by the modulator output filter and transmitting directional filter while the upper sideband is suppressed.

It is apparent from Fig. 16 that the received sideband coming in on the line from the distant repeater or terminal is operated upon by the filters in a reverse manner from that described above for the transmitted sideband. The incoming frequencies are directed through the receiving directional filter to the demodulator, where modulation with the original carrier reproduces the voice frequencies together with

other modulation products. The desired voice-frequency band is then separated from these products by the receiving low-pass filter.

In addition to performing the function of selecting desired and rejecting undesired currents, a filter, if operating in parallel with another as in the case of the directional filters or the line filters, should offer a high impedance to the transmitted currents of the other and thus prevent an excessive drain of these currents. Since these filters are

Fig. 17—Filter (with cover cut away) showing general method of assembly.

designed for operation in parallel, either filter operated without the other would have somewhat different electrical characteristics.

The economies and reduction in size of these filters as compared to those of the Type D system are due to several factors. Improved paper condensers having the required stability are used. Compared to mica condensers, these condensers are less costly and smaller for a given capacity. Moreover, at a very slight increase in cost paper condensers are used which will withstand 1000-volt line surges. Most of the condensers of the Type D system were designed for only 500-volt protection.

New types of coils with molybdenum permalloy cores having the necessary stability and low hysteresis losses are employed instead of

the air core solenoidal coils used in the Type D system. Since these new coils have less dissipation, the filters have flatter and wider transmission bands which contribute to improve telephone quality in the system. These coils are toroidal in shape and have very small stray fields and therefore small coupling with any nearby coils. This permits them to be packed closely together, which results in a large decrease in the size of the filters. A further reduction in size and cost of the filters was effected by dispensing with individual containers for each coil. The elements of the filter are wired together and held in place in the filter can by a potting compound. This is poured around them, hermetically sealing the whole assembly.

The small size of the elements and the very low coupling between them permit the assembly of more than one filter in the same can. For example, by a careful placing of the elements it was possible to place the transmitting and receiving low-pass filters and the modulator output filter in one can approximately $3\frac{1}{4}$ inches by $4\frac{1}{4}$ inches by $4\frac{1}{4}$ inches in size. A photograph of a high and low-pass line filter with the can cut away is shown in Fig. 17. The filters for a terminal of this system require 70 square inches of mounting space or about 1/5 of that required for those of a Type D system.

Conclusions

The development of the Type H system is another step in extending the use of carrier systems. Improvements in performance and simplifications which are effective in reducing its cost as compared with the Type D system which it supersedes have been obtained. Reduction in size and provision for operation on a-c. supply simplify its installation, particularly in outlying offices where suitable d-c. power supply is not ordinarily available. Its portability makes it well suited to provide additional circuits required in cases of emergency. The Type H system is expected to have a large application in the Bell System telephone plant, and in addition to provide carrier circuits for the communication systems of other companies.

Abstracts of Technical Articles from Bell System Sources

Directional Ferromagnetic Properties of Metals.[1] R. M. Bozorth. This is a review of the magnetic properties of single crystals of iron, cobalt and nickel and their alloys with each other. After a general introduction and a description of recently developed technique some new results are described which show for the first time that such crystals are anisotropic in low as well as in high fields. The amount of anisotropy is expressed usually by one "anisotropy constant" but sometimes a second constant is necessary. These constants vary with composition and temperature and their values are shown in tables and curves. When the constants are small the material is especially sensitive to strain and heat-treatment; this condition applies to some of the iron-nickel alloys (permalloys). Small amounts of impurities affect the magnetic properties in low fields. By careful purification and heat-treatment several crystals have been made which have permeabilities of over 1,000,000.

Use of Negative Regeneration in Radio Receivers.[2] C. B. Fisher. This paper gives a brief and simple explanation of gain stabilization which can be secured by employment of the feedback property of an amplifier.

Electron Diffraction Studies of Cuprous Oxide.[3] L. H. Germer. Etched surfaces of single crystals of cuprous oxide produce electron diffraction patterns consisting of spots and Kikuchi lines. In some cases lines are observed corresponding to all of the most widely spaced planes of copper atoms which are so situated as to be able to give reflections on the photographic plate. An example is given, however, of a diffraction pattern from which lines due to the most widely spaced plane of atoms (111), are missing when this plane lies parallel to the primary beam direction; the lines appear when the crystal is rotated to destroy this parallelism. The existence of Kikuchi line patterns indicates that the surfaces of etched cuprous oxide single crystals are relatively smooth. When spot patterns also appear they show anomalies which seem attributable to refraction, thus confirming the conclusion regarding smoothness. An example is given of an array of

[1] *Jour. of Applied Physics*, September 1937.
[2] *R. M. A. Engineer*, November 1937.
[3] *Phys. Rev.*, November 1, 1937.

spots made up of three distinct patterns which are displaced with respect to each other and with respect to the primary beam position. Polycrystalline surfaces made up of rather large crystals produce patterns of spots which are often drawn out into streaks lying approximately parallel to the plane of incidence. This effect, which is attributable to refraction and is a proof of smoothness of the surface, is most pronounced after etching in nitric acid. It is rarely, or never, found after etching in a mixture of hydrogen peroxide and ammonium hydroxide. Nitric acid etch produces upon the surface of a single crystal of cuprous oxide, or upon a polycrystalline surface made up of rather large crystals, crystals of some new compound which are cubic with cell edge 8.35A. These are sharply oriented with one or another of the most important planes parallel to the gross surface plane of the oxide. The sharpness of this orientation is proof of the flatness of the underlying oxide surface upon which the crystals are formed. This new 8.35A structure is produced by potassium cyanide etch as well as nitric acid, but never by sulphuric acid nor by ammonium hydroxide.

On the Ionization of the F_2 Region.[4] W. M. Goodall. In this paper the available data on F_2 region ionization for Peru, Australia, and this country are analyzed in a way that permits the separation of effects due to variations in solar ionizing force from effects due to seasonal and annual changes. It is shown that for constant solar activity the expected curves of critical frequency for Australia and this country appear to indicate both seasonal and annual tendencies. It is suggested as a possibility that the apparent "annual" effect may in fact be due to meteorological conditions which cannot be eliminated without data from more locations.

Coupling Between Parallel Earth-Return Circuits Under D-C Transient Conditions.[5] K. E. Gould. In tests conducted in connection with several d-c railway electrifications, the induced voltages recorded in paralleling communication circuits at times of short circuit on the railway have shown marked divergences from values computed on the basis of uniform earth resistivity and a rate of change of earth current determined from measurements in trolley and rail circuits. Due to the numerous factors which might contribute to these divergences, such as non-uniform division of transient current along the tracks and associated return conductors, the presence of shielding conductors along or near the right-of-way, etc., it was felt that a better understanding of the problem of induction under d-c transient conditions

[4] *Proc. I. R. E.*, November 1937.
[5] *Elec. Engg.*, September 1937.

could be obtained by experimental studies of the transient coupling between parallel earth-return circuits, free from the effects of shielding conductors, and with concentrated, rather than distributed, grounds. The study described in this paper was undertaken for this purpose.

Energy of Lattice Distortion in Cold Worked Permalloy.[6] F. E. Haworth. The lattice distortion produced by severe cold working of permalloy of 70 per cent Ni content has been studied by measuring the broadening of the reflection of the Fe $K\alpha$ doublet by the (311) planes, with a focusing camera. The broadening decreases upon annealing and recovery is complete at 650° C, when the breadth of the x-ray intensity curve at half-maximum is as small as that obtained by use of the two-crystal spectrometer. The mean square distortion in the lattice spacing due to cold work is derived from the X-ray measurements after photometering the x-ray film, converting the curve into an x-ray intensity curve, fitting the latter with an empirical equation and using an analysis worked out by S. O. Rice. The energy of the distortion is then calculated by using an equation derived by G. R. Stibitz. The root-mean-square distortion was found to be 0.31 per cent of the lattice spacing after the material had been reduced 96 per cent in cross-sectional area by cold working. The energy of distortion in the hard worked condition is thus found to be 23×10^6 ergs/cm^3 or 0.065 calorie/gram.

Optical Constants of Rubidium and Cesium.[7] Herbert E. Ives and H. B. Briggs. In previous papers the authors have presented the results of measurements of the optical constants of potassium and sodium in the ultra-violet and visible ranges of the spectrum, with a description of the apparatus used and the method of making measurements. The present paper deals with the results of measurements of the optical constants of rubidium and cesium by the same methods and for the same wave-length range. As with sodium and potassium, previous investigations of the optical constants of these metals were confined to the visible region of the spectrum.

The author's original interest in these constants lay in their application to the theory of photoelectric emission from thin films. In order to test this theory a knowledge of both the refractive index and the extinction coefficient of the metals concerned is needed for the spectral range embracing the characteristic photoelectric emission of the thin films. The work described in this paper is intended to supply this need.

[6] *Phys. Rev.*, September 15, 1937.
[7] *Jour. Opt. Soc. Amer.*, November 1937.

Minimum Noise Levels Obtained on Short-Wave Radio Receiving Systems.[8] KARL G. JANSKY. The theoretical minimum noise level of receivers in the absence of any interference, the source of which is external to the receiver, is discussed and compared with the limit actually measured on various antennas over a limited frequency range in the short-wave spectrum. It is pointed out that, on the shorter wave lengths and in the absence of man-made interference, the usable signal strength is generally limited by noise of interstellar origin. The powers obtained from this noise with the various antennas and for different times of the day are given.

Recently, man-made interference, of which that caused by diathermy machines constitutes the greatest part, has become so extensive that it is now the limiting noise during most of the daylight hours. Data are given on the intensity and extent of this form of interference.

Superstructures in Alloy Systems.[9] FOSTER C. NIX. A review of the literature treating superstructures in alloys. The presence of a superstructure produces new X-ray diffraction lines,—commonly called superstructure lines. The elements of superstructure theories are presented including the Bragg-Williams and Bethe-Peierls treatments. The author discusses the effect of a superstructure or an ordered phase on the thermal, mechanical, electrical and magnetic properties of alloys. A comparison is made between the theoretical predictions and the experimental results for both the specific heat and the energy of transformation. A Cu_3Au alloy becomes disordered more rapidly near the critical temperature of order, Tc, than the theories predict. A large anomalous specific heat is observed above Tc,—as predicted by the Bethe-Peierls theory.

Moment Recurrence Relations for Binomial, Poisson and Hypergeometric Frequency Distributions.[10] JOHN RIORDAN. This paper gives a uniform development of recurrence relations for moments about the origin and mean of binomial, Poisson, and hypergeometric frequency distributions. Uniformity is obtained through the use of the moment arrays of H. E. Soper. Both types of moments are expressed in terms of coefficients which are alike for the three distributions; for the moments about the origin these coefficients are the Stirling numbers of the second kind. Moment recurrence relations follow from recurrence relations for these coefficients. The recurrences for the hypergeometric moments appear to be new. For working purposes,

[8] *Proc. I. R. E.*, December 1937.
[9] *Jour. Applied Physics*, December 1937.
[10] *Annals of Math. Statistics*, June, 1937.

moments about the mean of binomial and Poisson distributions are expressed in terms of auxiliary moment coefficients with recurrence relations which also appear to be new.

Extending the Frequency Range of the Negative Grid Tube.[11] A. L. SAMUEL. The conventional vacuum tube when adapted for use at ultra-high frequencies carries with it many of the desirable attributes which it possesses at lower frequencies, such for example as its ability to amplify and the ease with which satisfactory frequency stability can be obtained. However, difficulties are encountered because of the effects of the finite electron transit time and because of certain circuit limitations. Methods of circumventing these restrictions are discussed and illustrated by reference to specific tubes. The paper closes with a brief review of recent work directed toward extending the frequency range of the negative grid tube both as an oscillator and as an amplifier.

Transmission Theory of Plane Electromagnetic Waves.[12] S. A. SCHELKUNOFF. This paper deals with transmission theory of plane electromagnetic waves in free space and in cylindrical regions of arbitrary cross section. Transmission properties of such waves can be expressed very simply in the same terms as the properties of electric waves guided by a pair of parallel wires. The earlier parts of the paper are concerned with general theorems and the latter parts with their application to plane waves in metal tubes of circular and rectangular cross sections.

The Empty Lattice Test of the Cellular Method in Solids.[13] W. SHOCKLEY. The cellular method of constructing wave functions for electrons in crystals developed principally by Wigner and Seitz and Slater is tested by applying it to an artificial crystal in which the potential is constant. Knowledge of the exact solutions for this case, plane waves, shows that the cellular method is quite accurate in the first Brillouin zone but may be in error by a factor of two in the second. Hence calculations of occupied levels in Li and Na are probably quite good; for Cu, Ca, diamond, LiF, and NaCl the errors will be larger. Calculations of excited states are likely to be very much in error. The accuracy of the cellular method is shown to improve very slowly with increasing number of continuity conditions.

Sound Propagation in Ducts Lined with Absorbing Materials.[14] L. J. SIVIAN. In ventilator and exhaust systems it is desirable to provide

[11] *Jour. of Applied Physics*, October 1937.
[12] *Proc. I. R. E.*, November 1937.
[13] *Phys. Rev.*, October 15, 1937.
[14] *Jour. Acous. Soc. Amer.*, October 1937.

a high degree of attenuation for audio-frequency waves while offering low resistance to continuous or slowly pulsating air flow. For that purpose ducts lined with absorbing materials are sometimes used. This paper deals with sound propagation, particularly with its attenuation constant, in such ducts. Two types of lining are considered: (a) non-vibratile, i.e., a lining in which there is no wave motion propagated in the direction of the duct axis; (b) vibratile, i.e., a lining admitting of such motion. Methods for computing the propagation constants in terms of the acoustic constants of the lining, are given. Some experimental data are presented. Comparison of observed and computed values indicates that the computational procedure is substantially valid up to a frequency at which the sound wave-length is about twice the internal duct diameter.

New Experimental Methods Applicable to Ultra-Short Waves.[15] G. C. SOUTHWORTH. This paper presents a new approach to the problem of electrical measurements at extremely high frequencies. It makes use of a relatively new principle whereby electromagnetic waves may be propagated through the interiors of hollow metal tubes. These tubes have for convenience been called wave guides. Their diameters must be at least 0.585 wave-length in order that power may be propagated. Some of the difficulties of previous methods have been incidental to radiation and spurious coupling effects. These have been largely eliminated by this method since the power resides almost entirely within the pipe. Short sections of wave guide may be made to resonate electrically much as organ pipes and air columns resonate acoustically. A high degree of sharpness may be obtained. Such a pipe may, therefore, play the role of a resonant circuit and become effectively a wave meter, a frequency determining unit for an oscillating vacuum tube or an impedance matching device. Specimens under study may be placed inside a resonant chamber and be subjected to electric fields of controlled intensities.

Preparation of Large Single Crystals of Sodium Chloride.[16] H. WALTHER. An apparatus and a method are described, whereby single crystals of sodium chloride are produced in the form of bars 2 cm in diameter and 30 cm long. The crystal is drawn from the melt by means of a platinum rod which is dipped into it and which is raised and rotated simultaneously by a clock mechanism. An air stream from a circular nozzle surrounding the growing crystal immediately above the surface of the molten salt provides the steep

[15] *Jour. Applied Phys.* October 1937.
[16] *Rev. Sci. Instruments*, November 1937.

temperature gradient necessary for the continuous growth of the crystal. The orientation of the crystal with respect to the axis of the bar may be chosen at will.

Magnetic Properties of Single Crystals of Silicon Iron.[17] H. J. WILLIAMS. The magnetization curves for the [100], [110] and [111] directions of single crystals of iron containing 3.85 per cent silicon were obtained from single crystal specimens cut in the form of hollow parallelograms so that the sides of each specimen were parallel to the tetragonal, digonal or trigonal axes, respectively. This method avoids the errors due to demagnetizing fields, inherent in previous measurements on single crystals. In addition to the well-known anisotropy at magnetizations above half of saturation, the data show for the first time considerable anisotropy at low magnetizations. A maximum permeability of 1,380,000, by far the highest ever reported for silicon iron, was observed in the [100] specimen after careful annealing. The magnetic anisotropy constants K_1 and K_2 were obtained from the magnetization curves and from torque measurements on a disk cut parallel to a (110) plane.

The Quest of Vitamin B_1.[18] R. R. WILLIAMS. An account is given in semi-popular form of the events which led to a recognition of a dietary deficiency as the cause of Oriental beriberi. The steps are indicated which led to the isolation and synthesis of the lacking substance which is now known as vitamin B_1. The author began his researches on this subject in Manila in 1910 and has prosecuted them continuously since that time. In 1936 he had the gratification of effecting with Dr. J. K. Cline of Merck and Company a practical synthesis of this vital compound so that it is now produced artificially much more cheaply than it can be obtained in pure form from nature.

In presenting his work on the architecture of the molecule and its artificial reconstruction from synthetic sources, the author endeavors to make clear to the layman by analogy and otherwise the methods which the organic chemist employs in such a study. The highly inferential deductions concerning complex interatomic relationships which result from tearing the natural molecule apart reach a dramatic verification only when one finds he has actually reproduced in the laboratory the precise product of nature's art.

[17] *Phys. Rev.*, October 1, 1937.
[18] *Jour. Franklin Institute*, November 1937.

Contributors to this Issue

M. L. Almquist, B.S. in Electrical Engineering, University of California, 1920. General Electric Company, 1920–21. American Telephone and Telegraph Company, Department of Development and Research, 1921–1934; Bell Telephone Laboratories, 1934–. Mr. Almquist has been engaged in transmission development work on carrier telephone systems.

F. H. Best, M.E., Cornell University, 1911. American Telephone and Telegraph Company, Engineering Department, and Department of Development and Research, 1911–34. Bell Telephone Laboratories, 1934–. Mr. Best has been engaged principally in the development of methods and arrangements used in maintaining the transmission efficiency of telephone circuits.

R. W. Chesnut, A.B., Harvard University, 1917; War Department of French Government, 1917; U. S. Army, 1917–19. Western Electric Company, Engineering Department, 1920–25; Bell Telephone Laboratories, 1925–. Mr. Chesnut has been engaged in the development of carrier telephone and long-wave radio systems.

Paul S. Darnell, B.S. in Electrical Engineering, University of Pennsylvania, 1922; M.A., Columbia University, 1925. Western Electric Company, Engineering Department, 1922–1925; Bell Telephone Laboratories, 1925–. Mr. Darnell's work has been primarily in connection with the development of retardation coils for operation at voice and carrier frequencies.

Clifford E. Fay, B.S. in Electrical Engineering, Washington University, 1925; M.S., 1927. Bell Telephone Laboratories, 1927–. Mr. Fay has been engaged principally in the development of power vacuum tubes for radio purposes.

H. J. Fisher, E.E., Cornell University, 1920; U. S. Signal Corps, 1917–19. Western Electric Company, Engineering Department, 1920–25; Bell Telephone Laboratories, 1925–. Mr. Fisher has been engaged in the development of toll switching and transmission systems.

C. W. Green, B.S. in Electrical Engineering, University of Wisconsin, 1907; Instructor and Assistant Professor, Massachusetts Institute

of Technology, 1907–17; Captain 1917, Major 1918, U. S. Army. Bell Telephone Laboratories, 1919. Mr. Green's work has had to do with the development of carrier telephone systems, voice frequency repeaters, and radio terminating circuits.

E. I. Green, A.B., Westminster College, Fulton, Missouri, 1915; University of Chicago, 1915–16; Member of Faculty, Westminster College, 1916–17; U. S. Army, 1917–19 (Captain, Infantry); B.S. in Electrical Engineering, Harvard University, 1921. American Telephone and Telegraph Company, Department of Development and Research, 1921–34; Bell Telephone Laboratories, 1934–. Mr. Green's work has been concerned principally with multiplex transmission systems.

L. M. Ilgenfritz, B.S. in Electrical Engineering, University of Michigan, 1920. American Telephone and Telegraph Company, Department of Development and Research, 1920–34; Bell Telephone Laboratories, 1934–. Mr. Ilgenfritz has been engaged in the development of carrier systems.

A. Kenner, B.S. in Electrical Engineering, Purdue University, 1913. Western Electric Company, Engineering Department, 1913–25; Bell Telephone Laboratories, 1925–. Mr. Kenner has been engaged in equipment development work for toll systems, particularly those items such as telephone repeaters and carrier associated with toll cables.

C. E. Lane, A.B., University of Iowa, 1920; M.S., University of Iowa, 1921. Western Electric Company, Engineering Department, 1921–25; Bell Telephone Laboratories, 1925–. Mr. Lane is engaged in the development of special filters such as mechanical filters, filters in which quartz crystals are used as elements, and high-frequency filters of all types.

R. H. Mills, S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1916. Western Union Telegraph Company, 1916–18. Western Electric Company, Transmission Development Branch, 1918–25; Bell Telephone Laboratories, Apparatus Development Department, 1925–. Mr. Mills has been engaged in the development of carrier frequency filters for use in commercial communication systems.

A. L. Samuel, A.B., College of Emporia (Kansas), 1923; S.B. and S.M. in Electrical Engineering, Massachusetts Institute of Technology,

1926. Instructor in Electrical Engineering, Massachusetts Institute of Technology, 1926–28. Bell Telephone Laboratories, 1928–. Mr. Samuel has been engaged in research and development work on vacuum tubes.

S. A. Schelkunoff, B.A., M.A. in Mathematics, The State College of Washington, 1923; Ph.D. in Mathematics, Columbia University, 1928. Engineering Department, Western Electric Company, 1923–25; Bell Telephone Laboratories, 1925–26. Department of Mathematics, State College of Washington, 1926–29. Bell Telephone Laboratories, 1929–. Dr. Schelkunoff has been engaged in mathematical research, especially in the field of electromagnetic theory.

W. Shockley, B.Sc., California Institute of Technology, 1932; Ph.D., Massachusetts Institute of Technology, 1936. Bell Telephone Laboratories, 1936–. Dr. Shockley's work in the Laboratories has been concerned with problems in electronics.

Rexford S. Tucker, A.B., Harvard University, 1918; S.B. in Electrical Engineering, Harvard Engineering School, 1922. American Telephone and Telegraph Company, Department of Development and Research, 1923–1934. Bell Telephone Laboratories, 1934–. Mr. Tucker's work has been on noise and crosstalk problems.

M. A. Weaver, E.E., Lehigh University, 1915. American Telephone and Telegraph Company, Long Lines Department and Department of Development and Research, 1915–34; Bell Telephone Laboratories, 1934–. Mr. Weaver is engaged in development work connected with crosstalk and noise problems of voice and carrier-frequency circuits.

The Bell System Technical Journal

Vol. XVII *April, 1938* *No. 2*

Studies of Telephone Line Wire Spacing Problems

By J. A. CARR and F. V. HASKELL

By spacing a wire and its mate closer than has been the usual custom, and thus obtaining a greater separation between pairs on open wire telephone lines, inductive interference can be diminished and higher carrier frequencies transmitted, thus increasing the message carrying capacity. The hazard of the wires swinging into contact in the wind is controlling in the selection of a minimum spacing. This article gives the results of studies made to determine the effect of natural winds on variously spaced and arranged wires suspended in accordance with telephone practice.

UNTIL recent years, it has been the usual practice in the Bell System to use a spacing of 12 inches between adjacent wires on a crossarm with the exception of the wires of the pole pair which are more widely separated to provide climbing space. By reducing this 12-inch spacing between the wires of non-pole pairs and obtaining a correspondingly greater separation of pairs, a decrease is obtained in the crosstalk coupling between pairs and also in the induced voltage from such external sources as radio stations and static. This permits the transmission of higher carrier frequencies and provides a greater number of communication channels without impairing the quality of transmission. The increased number of channels that can be obtained is a function, among other factors, of the magnitude of the change in wire spacing.

It was with a view to ascertaining the limitations that might be imposed upon such a rearrangement of the wires through their swinging into contact in winds that the study herein described was initiated by Bell Telephone Laboratories.

It is interesting to note here that, since the early results of this study became available, about 75,000 pair-miles of toll line wire have been installed with an 8-inch spacing between the wires of a pair and recently some 2,600 pair-miles with a 6-inch spacing between the wires of a pair.

At the outset of this study, considerable thought was given to a theoretical determination of the limiting factor or the chance of two parallel wires contacting in the wind, but with little success. Difficulty was encountered in taking into consideration some of the factors involved, principally the gusty nature of the winds to which a pole line is frequently subjected. There was evolved out of this attack, however, a theory embodying the movements of a single suspended wire in wind. This theory, which was presented in a previous article,[1] contributed to the general knowledge of wire movements in wind and gave information on the equilibrium position of a span of wire in a steady wind and the movements of the wire about this position with gusts of wind superimposed on this steady wind. This theory is referred to later.

As important as this theory is, it did not solve the problem of determining the chances of two parallel wires swinging together in natural winds. Consideration was also given to the possibility of obtaining this information through tests of model lines in a wind tunnel. While the problem of meeting the requirements of dynamic similarity in these model tests did not appear to offer serious trouble, the problem of simulating natural wind conditions did present difficulties of major proportions. The line of approach finally adopted was a full scale test of variously spaced and arranged wires on lines erected at a location especially selected for its unusual exposure to natural winds.

This site which is in the township of Chester, New Jersey, is about 950 feet above mean sea level. It is the highest point for many miles in every direction and is particularly well exposed to the prevailing winds which are northwest in this locality.

Provision of Test Facilities

On this site there was first erected a six span line, referred to herein as Line 1, in a direction as nearly perpendicular to the prevailing northwest winds as the contour of the ground would permit without excessive pole heights at the extremities. The span between poles was 130 feet, the Bell System standard length for heavy open wire lines, except at the ends where two poles were set a few feet apart in line for convenience in testing. Figure 1 gives a view of the line and Fig. 2 the special end construction. This arrangement made it possible to include span lengths of 130 and 260 feet in any test. The latter span length was obtained by attaching the wires to the lower

[1] "Motion of Telephone Wires in Wind," D. A. Quarles, *Bell System Technical Journal*, April 1930.

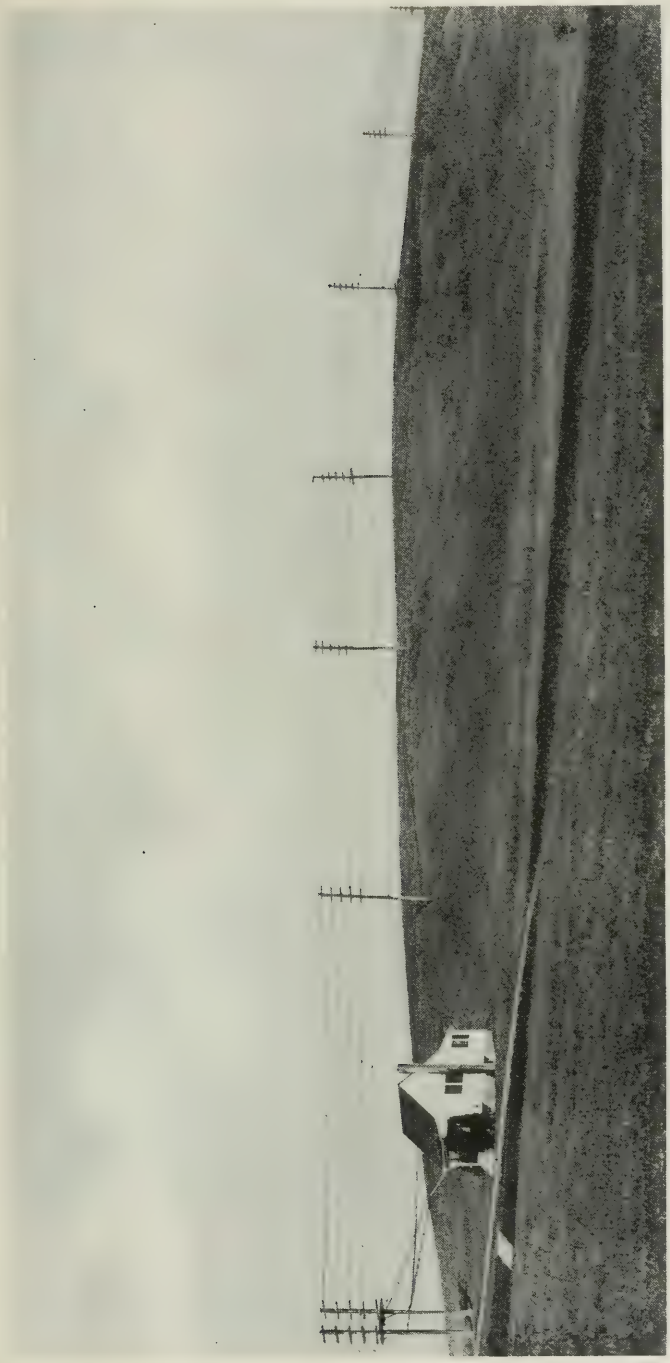


Fig. 1—Line 1 and test house at Chester, New Jersey.

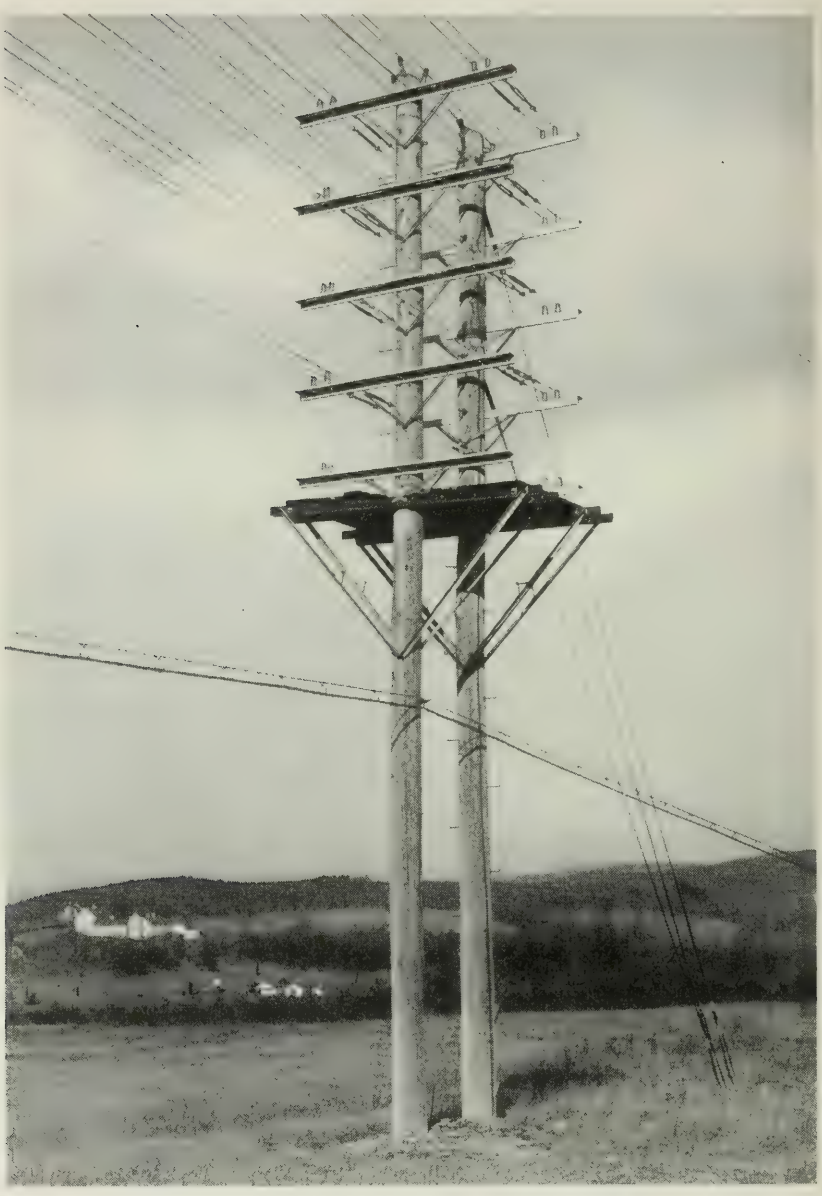


Fig. 2—Double pole construction at end of Line 1.

crossarms at alternate poles. Later another line (Line 2) was erected. This line consisted of three spans, one each of 100, 160 and 260 feet.

Special steel crossarms were mounted on each pole with a vertical separation between them somewhat greater than the two feet usually used in the Bell System. This was to lessen the effect of wind shielding on the test results. These steel crossarms were slotted so as to accommodate steel insulator pins and make it possible to vary the spacing at will. On each pole of Line 1 were mounted five of these crossarms each carrying eight pins, two pairs on each side of the pole. The line wires used were hard drawn copper, 0.104 inch and 0.165 inch in diameter. These sizes comprise the two extremes of the three sizes of copper line wire most frequently used in the Bell System, the intermediate size being 0.128 inch in diameter. As the results obtained during the first two years indicated that the size of wire, within the range of interest, did not appreciably affect the contacting tendency, later tests were confined to the 0.165-inch diameter wire which was selected because of its relatively greater use for important circuits. On Line 2, one crossarm with four pairs of 0.165-inch diameter wire was mounted in the 260-foot span and two crossarms of the same size wire in the 100- and 160-foot spans.

The spacings used in the tests between the wires of a pair comprised 3, 4, 6, 8 and 12 inches, a range which starts with as small a horizontal spacing as appeared to offer possibilities, with the present type of construction, and extends up to the previous standard of 12 inches. The sags used conformed to Bell System Practices and ranged from 4 inches to 45 inches depending upon temperature and span length. In certain of the tests sag inequalities between the wires of a pair were introduced to simulate chance conditions. In referring to the results unless otherwise noted it is to be understood that the two wires of a pair were at equal sags. When the sags in the wires were unequal it is so stated. The amount of the inequality used in spans of 100, 130 and 160 feet was 3 inches at 60° F., while that used in 260-foot spans was 6 inches at 60° F. In these tests the greater sag was always placed in the windward wire while the sag in the leeward wire was maintained so as to conform to Bell System Practices.

In the later stages of the tests certain anti-contacting devices as hereafter described were applied to the wires in the narrow spaced arrangements to raise the velocities at which contacting began.

Recording Apparatus

Apparatus was installed in a building adjacent to one end of the line to record graphically the number and time of occurrence of the

swinging contacts between the wires of each pair under test and simultaneously to record the velocity and direction of the wind.

To obtain an idea of the nature of a contact, oscillograph studies were conducted at the outset. The oscillograms (Fig. 3) are repre-

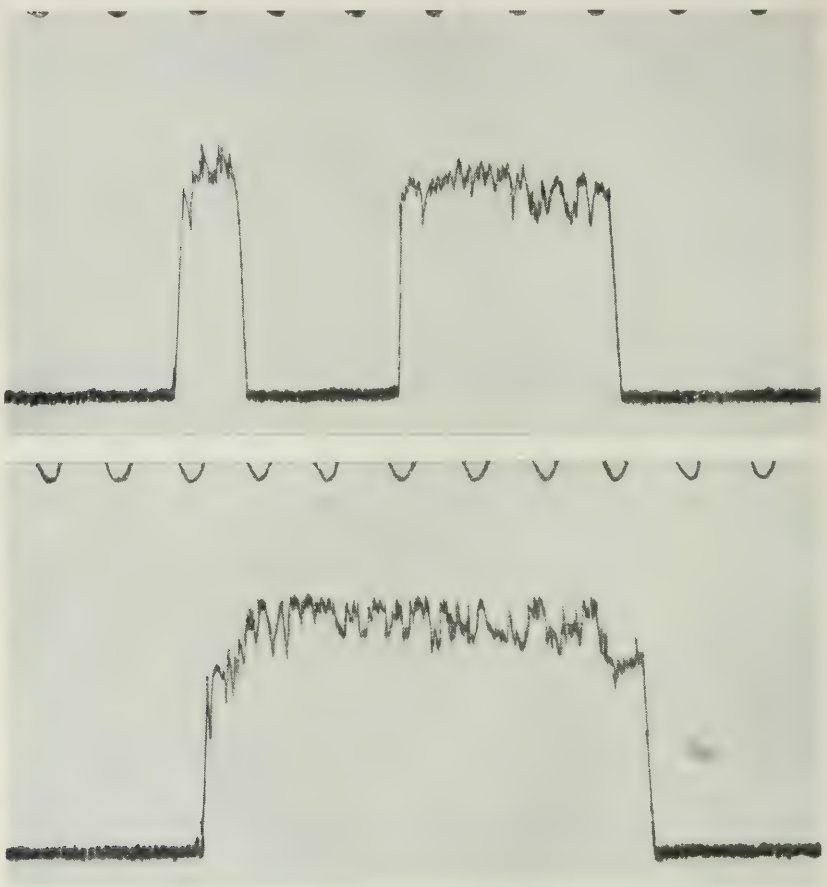


Fig. 3—Oscillograms of swinging wire contacts.

sentative of the results obtained. The timing wave has a frequency of 60 cycles. When swinging into contact, it appears that the wires scrape and chatter as evidenced by the rapid variations in the current flow. It was found in these oscillograph studies that during a single contact the resistance might vary from zero to 50,000 ohms while in another case the variation might be from zero to 5,000 ohms. The duration of a contact varied approximately from 0.004 to 0.230 second.

However, the recording apparatus was designed to record a contact that would affect telegraph transmission, which mode of communication was considered to be more sensitive to this type of interference than any other form of transmitted signal. This contact was defined as having a minimum duration of 0.01 second and a resistance of 20,000 ohms or less. The voltage applied to the lines during the test was 260. This voltage was the highest that would obtain on an open wire d-c. telegraph line and would occur when one side of the circuit had a potential of minus 130 volts and the other side plus 130 volts.

Considerable study was given to the selection of apparatus for recording wind velocities. After studying the various types of available instruments and consulting with the United States Weather Bureau, a Burton type of anemometer associated with an Esterline-Angus graphic recorder was selected as best meeting our requirements. This anemometer assembly comprised a four-cup anemometer with the armature of a magneto mounted on the cup-shaft. The magneto current which varies with the rotational speed of the cups was recorded on a milliammeter specially designed for the purpose. This recording milliammeter had several chart speeds so that considerable detail in the velocity change could be obtained when required. The instrument was calibrated to read directly the instantaneous velocity of the wind. A wind vane which registered sixteen directions of wind provided a continuous record of wind direction on the same chart with the wind velocity. The mounted anemometer and the wind vane are shown in Figs. 4 and 5, respectively.

It was realized in selecting such a velocity recorder that our records would not be directly comparable to the weather bureau records which are five-minute average velocities rather than instantaneous velocities unless relations between the two could be established. For this reason when analyzing the wind velocity charts two velocities were read, namely the average and the maximum for each five-minute interval. To further the comparison it was established that the maximum instantaneous velocity ranged greater than the five-minute average velocity by an average figure of 1.4 for velocities greater than 15 miles per hour. This figure which applies to the Chester location and the Burton instrument was obtained by taking the average ratio of the maximum velocities to the five-minute average velocities for 500 cases. Figure 6 shows the frequency distribution of these 500 cases. This graph indicates that the modal value of the ratio as well as the arithmetic mean is approximately equal to 1.40. Two-thirds of these cases fall between ratio values of 1.25 and 1.55.

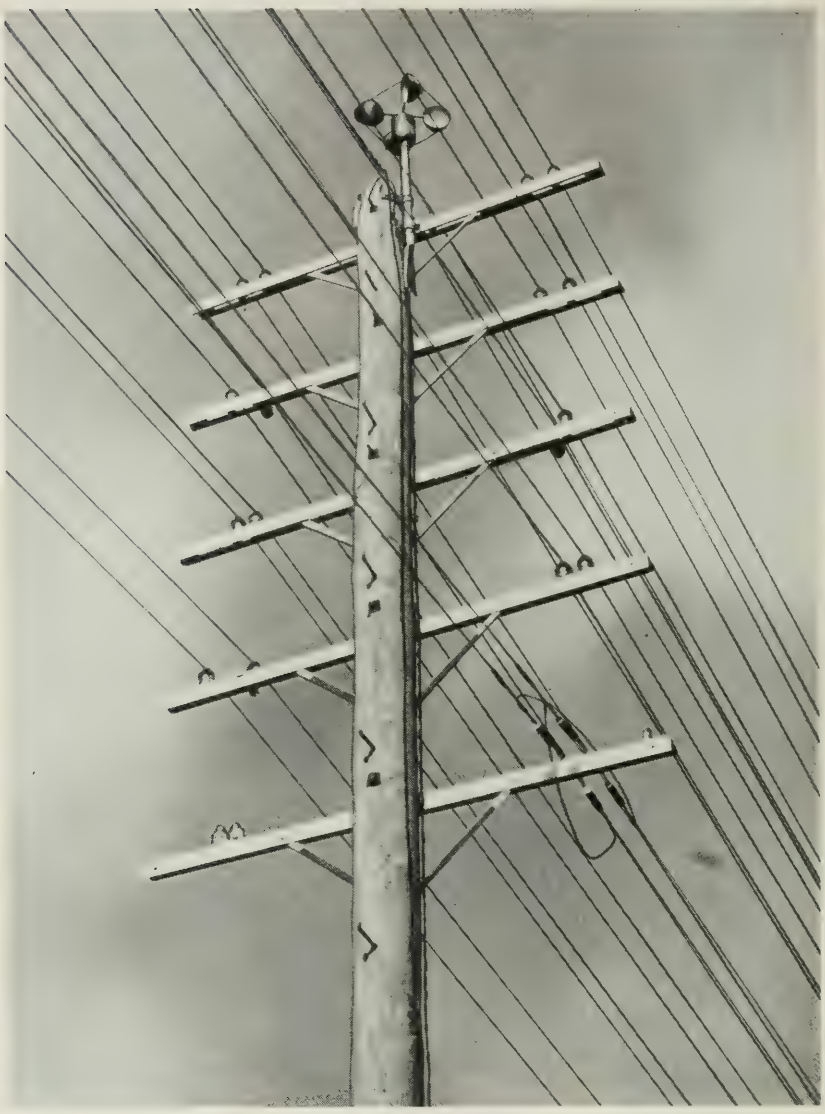


Fig. 4—Cup anemometer mounted on pole of Line 1.

Fig. 5—Wind vane mounted on pole of Line 1.

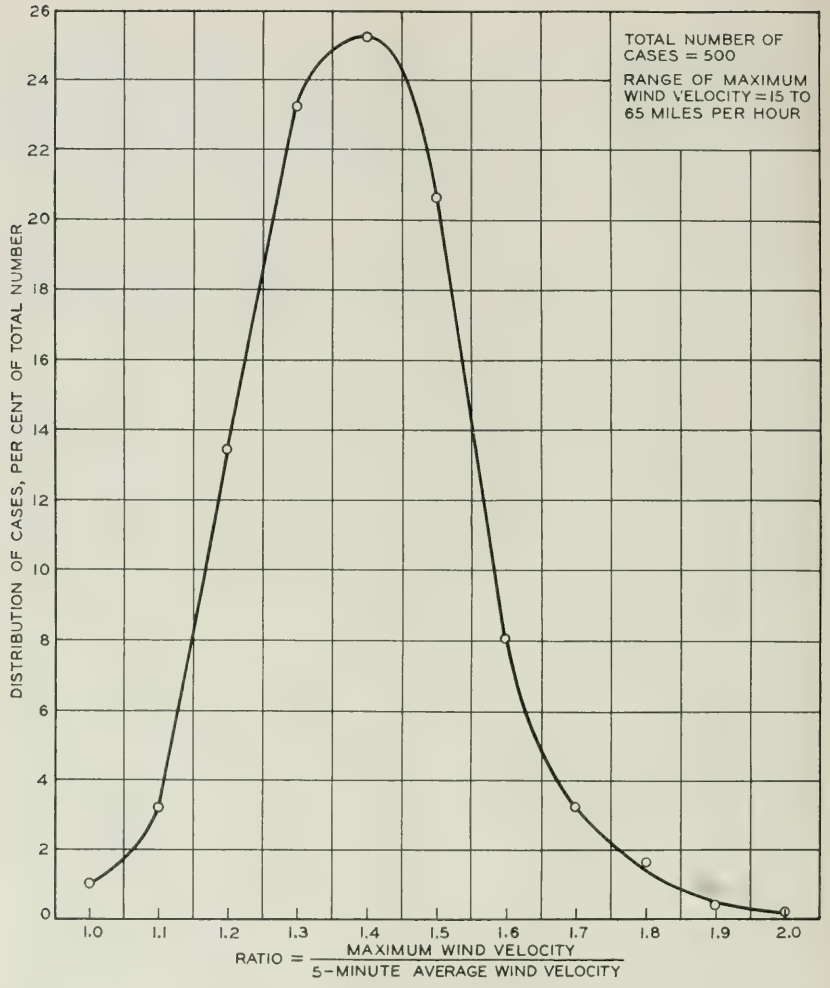


Fig. 6—Frequency distribution of the ratios of maximum instantaneous velocities to five-minute average velocities as recorded by the Burton anemometer.

Method of Treating Data

The data considered necessary for the characterization of any pair arrangement were:

1. The number of contacts.
2. The simultaneous velocity and direction of the wind.
3. The sags of the wires.
4. The air temperature (because of its effect on wire sag).
5. The weather, particularly the presence or absence of glaze.

To simplify to some extent the long process of correlation and analysis, first, the velocities recorded at 16 directions to the line were converted to normal velocities, which should be borne in mind in considering the results. This conversion was made by multiplying the recorded velocities by the cosine of the angle between the normal to the line and the true direction of the wind. The propriety of taking this step was based on appropriate wind tunnel tests made at New York University and reported in the *Bell System Technical Journal.*[2] From this point on the data were classified according to the procedure outlined in Fig. 7.

This procedure provided a history of the contacting on each pair arrangement tested. This history detailed the number of contacts occurring in each five-mile-per-hour cell of maximum and five-minute average normal wind velocity for each division of sag. Each division of sag comprised a cell of one to three inches depending upon the length of span.

These results were analyzed with the view of determining first, the instantaneous velocity, at which contacting begins to occur for each pair arrangement, and second, any relationships existing between the fundamental factors of spacing, sag, span length and such instantaneous velocities which are hereinafter referred to as threshold velocities. The term "threshold velocity" as used in this article does not relate to the five-minute average velocities but to the maximum velocities previously mentioned. To express threshold velocities in terms of five-minute average velocities it is necessary to divide by the factor of 1.4, referred to above.

The analysis of the data directed towards determining the first objective or the threshold velocity for each pair arrangement revealed considerable variation in the magnitudes of these velocities for any particular sag. This has been ascribed to the variability of natural winds. Under these conditions it appeared to be appropriate to select the velocity most frequently associated with the beginning of contacting as the threshold velocity for a given arrangement at the prevailing sag. Thus the modal value was taken as the threshold velocity, or the nearest velocity when expressed in multiples of five miles per hour to the true modal value. The accompanying Table I lists the threshold velocities for the arrangements tested.

Regarding the second objective, namely, the analysis of the data for the purpose of determining relationships existing between the fundamental factors of spacing, span length, sag and wind velocity,

[2] "Forces of Oblique Winds on Telephone Wires," J. A. Carr, *Bell System Technical Journal*, October 1936.

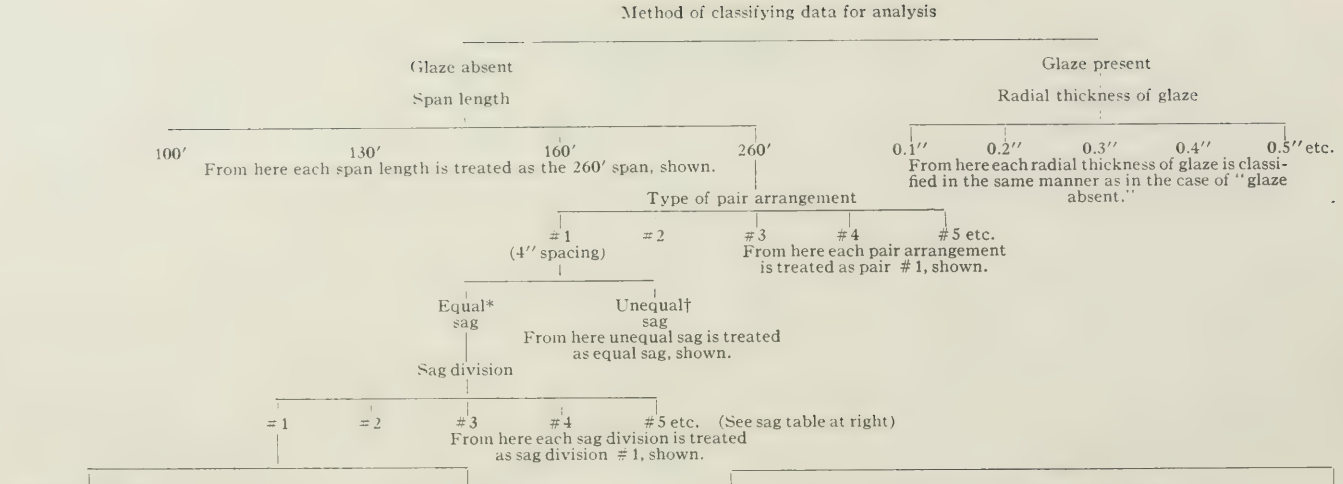

Method of classifying data for analysis
Glaze absent
Span length
100′
130′
160′
260′
From here each span length is treated as the 260′ span, shown.
Glaze present
Radial thickness of glaze
0.1″
0.2″
0.3″
0.4″
0.5″ etc.
From here each radial thickness of glaze is classified in the same manner as in the case of "glaze absent."
Type of pair arrangement
#1 (4″ spacing)
#2
#3
#4
#5 etc.
From here each pair arrangement is treated as pair #1, shown.
Equal* sag
Unequal† sag
From here unequal sag is treated as equal sag, shown.
Sag division
#1
#2
#3
#4
#5 etc. (See sag table at right)
From here each sag division is treated as sag division #1, shown.
Maximum normal wind velocity
Number of 5-minute periods the wind is in each velocity cell (0 to 5, 6 to 10, etc. m.p.h.) and number of contacts in each period.
5-minute average normal wind velocity
This is treated the same as the maximum normal wind velocity.

Table of sags (inches) from Bell System Practices G31.115

Sag division	1	2	3	4	5	6	7	8	9	10	11
Temp. ° F.	0	10	20	30	40	50	60	70	80	90	100
Span (feet) 100	3.5	3.5	4	4	5	5.5	6	6.5	7.5	8.5	9.5
130	5.5	6	6.5	7	8.5	9	10	11.5	12.5	14	16
160	9	9.5	10.5	11.5	12.5	14	15	16	18	19.5	21
260	25.5	27.5	30	32	34.5	37.5	40	43	46	49	52

* Equal sag—Both wires of the pair were maintained at the sag given in Bell System Practices G31.115.

† Unequal sag—The wire away from the prevailing winds was maintained at the sag given in Bell System Practices G31.115 while the windward wire had a greater sag by an amount based on a figure of 3 inches at 60° F. for the 100, the 130, and the 160-foot spans and 6 inches at 60° F. for the 260-foot span.

Fig. 7—Manner in which data were classified for analysis.

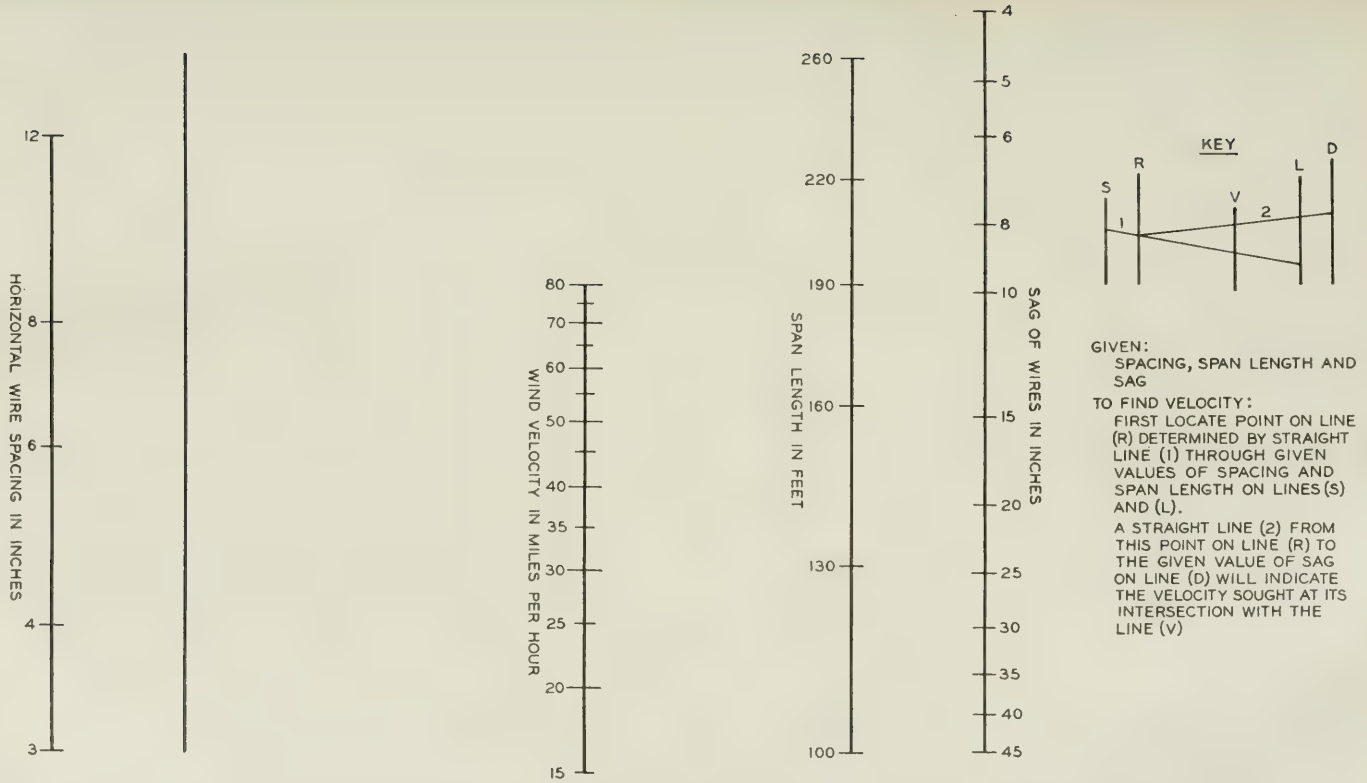


Fig. 8—Nomogram for determining wind velocity (threshold) normal to the line at which wires of a pair begin to contact.

only a moderate degree of success was obtained. From an examination of the data in Table I, it appears for the condition "Equal Sag" that threshold velocities increase with the spacing between the wires and the span length and decrease with an increase in the sag. Because large sags usually accompany long spans it may be somewhat difficult to appreciate the fact that the threshold velocities increase with span length. This is a fact, however, when sag and wire spacing remain constant. Using this basic relation an empirical equation (5) has been developed from the data obtained and is given in Appendix II. A nomogram was constructed for this equation (5) and is given in Fig. 8.

Results of Natural Wind Tests

In Fig. 9 are given two curves of the wind velocities recorded at Chester, New Jersey, during the first seven of the eight years the test was in progress. One of the curves (marked, "Maximum") gives the average annual frequency of occurrence in terms of five-minute periods of maximum wind velocities grouped in cells of five miles per hour. The other curve gives the corresponding data for the five-minute average wind velocities. The maximum velocity reached 60 miles per hour on several occasions each year and exceeded 70 miles per hour on at least one occasion. The velocities occurring during the test are considered to be as great as or greater than those to which the structural plant is usually subjected.

Regarding the data obtained during the presence of glaze, wires with the spacings and sags tested contacted at velocities as low as 10 to 15 miles per hour. To some extent the number of contacts increased with the amount of glaze and the velocity of the wind. Also, generally, a greater number of contacts occurred on the more closely spaced wires than on those with greater spacings. In cases of wires spaced 3 and 4 inches apart there were occasions when they contacted and froze together for periods of several hours. Due to the erratic action of the wires during the presence of glaze not all of these data were classified in detail as were the results obtained when glaze was absent.

The classified data obtained during the absence of glaze in terms of normal threshold wind velocities (maximum) are given in Table I for all the arrangements tested, approximately eighty-five. The data given in this table were collected over a period of approximately eight years. While none of the arrangements were under test over the whole period, all were under test for at least one winter season during these eight years. Since the higher velocities occurred more frequently in the winter than in the summer most of the threshold velocity data

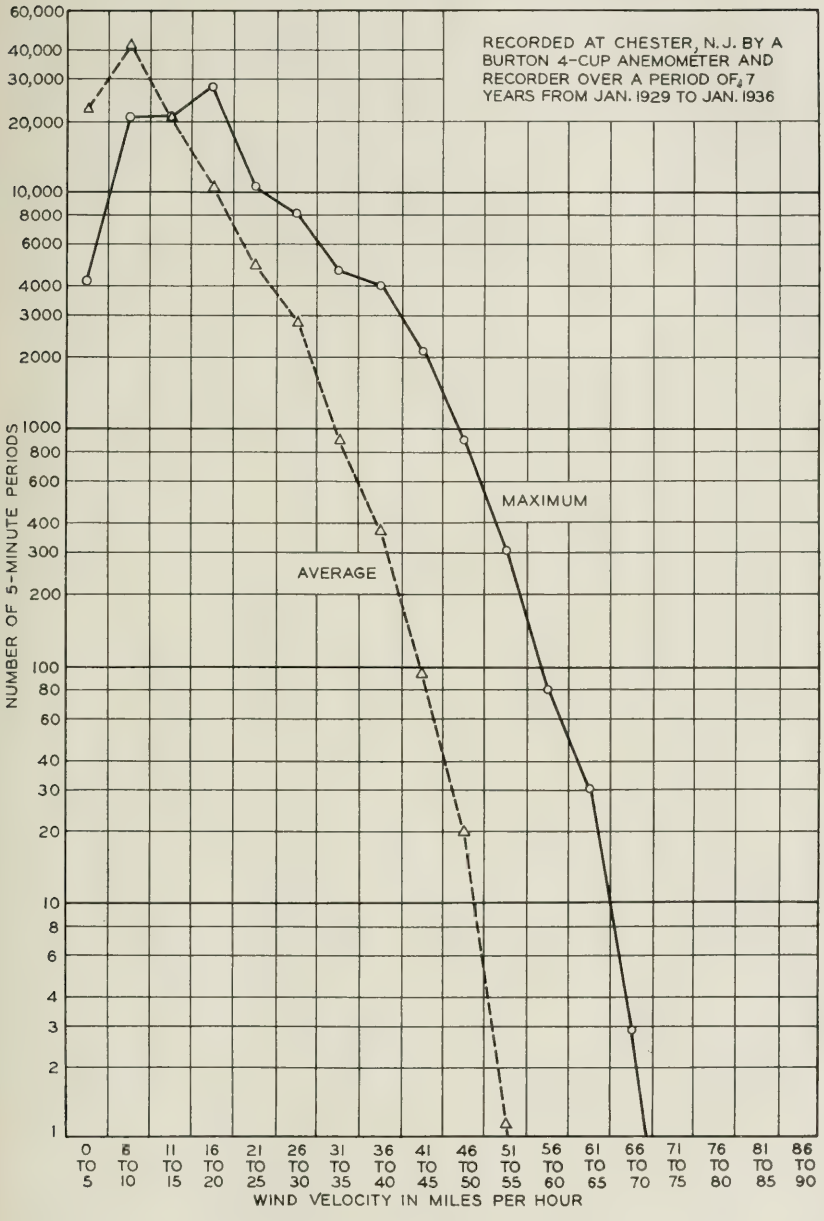


Fig. 9—Distribution of the annual maximum and five-minute average wind velocities.

TABLE I

Wind Velocity * (Threshold) Normal to the Line at Which Wires of a Pair Begin to Contact **
Jan. 1, 1929 to May 1, 1936

Equal Sag [2]

100-Foot Span [1]

Type of Circuit	Temp. °F. 20	30	40	50	60	70	80	90
Sag—Inches	4.0	4.0	5.0	5.5	6.0	6.5	7.5	8.5
3″ Spacing, Unequipped Wires		>60		50	50		45	40
4″ " " "		>60		>60	60		>55	45
6″ " " "		>60		>60	—		—	—
8″ " " "		>60		>60	—		—	—
3″ Spacing, 1–3″ Diam. Disc				—	55	—		
3″ " 1–4″ " "				>65	—	—		
4″ " 1–4″ " "				>65	—	65		

130-Foot Span [1]

Type of Circuit	Temp. °F. 20	30	40	50	60	70	80	90	100
Sag—Inches	6.5	7.0	8.5	9.0	10.0	11.5	12.5	14.0	16.0
3″ Spacing, Unequipped Wires	40		40	40	—	30	30	30	—
4″ " " "	50		45	40	40	35	—	35	30
6″ " " "	70		>60	>60	55	50	45	45	40
8″ " " "	>70		—	—	—	—	>70	—	>70
12″ " " "	—		—	—	>70	—	—	>70	—
3″ Spacing, 1–3″ Diam. Disc	40				35			35	
" " 2–3″ " "	—				40			40	
" " 3–3″ " "	50				40			40	
" " 1–4″ " "	55	55		45	40	40		40	
" " 2–4″ " "	—	—		—	>65	—		>45	
" " 3–4″ " "	—	—		—	>50	—		>45	
4″ " 1–3″ " "	50				45			45	
" " 2–3″ " "	—				50			>45	
" " 3–3″ " "	65				65			>45	
" " 1–4″ " "	55	55	50	45	45		45	45	
" " 2–4″ " "	—	—	—	—	50		—	>45	
" " 3–4″ " "	—	—	—	—	>65		—	>45	
6″ " 1–4″ " "								60	

Unequal Sag [2]

100-Foot Span [1]

Type of Circuit	Temp. °F. 20	30	40	50	60	70
Sag—Inches	4.0	4.0	5.0	5.5	6.0	6.5
3″ Spacing, Unequipped Wires	40	35	—			35
4″ " " "	45	40	40			40
3″ Spacing, 1–4″ Diam. Disc			55			
4″ " 1–4″ " "			60			

130-Foot Span [1]

Type of Circuit	Temp. °F. 20	30	40	50	60	70	80	90	100
Sag—Inches	6.5	7.0	8.5	9.0	10.0	11.5	12.5	14.0	16.0
3″ Spacing, Unequipped Wires	—	35	—		35		35		
4″ " " "	—	40	—		35		35		
6″ " " "	>50	45	45		—		—		
8″ " " "	—	>65	—		>55		—		
12″ " " "	—	>65	—		>55		—		
3″ Spacing, 1–3″ Diam. Disc		45			—		—		
" " 2–3″ " "		45			45		—		
" " 3–3″ " "		50			45		40		
" " 1–4″ " "		50			45	—	40	30	
" " 2–4″ " "		50			50	—	—	—	
" " 3–4″ " "		55			—	>50	—	—	
4″ " 1–3″ " "		45			—				
" " 2–3″ " "		45			—				
" " 3–3″ " "		50			—				
" " 1–4″ " "		50			40	—	—	40	
" " 2–4″ " "		50			50	—	50	—	
" " 3–4″ " "		>45			—	>50	45	40	
6″ " 1–4″ " "	>50		55					55	

100-Foot Span [1]

	Temp. °F. 20	30	40	50	60	70	80
	Sag— Inches 10.5	11.5	12.5	14.0	15.0	16.0	18.0
3″ Spacing, Unequipped Wires	40	40	35	30	—	—	
4″ " " "	45	45	40	35	35	—	
6″ " " "	55	50	45	—	—	—	
8″ " " "	—	>55	—	—	—	>55	
3″ Spacing, 1–3″ Diam. Disc	—	—	—		40		
" " 3–3″ " "	—	—	—		60		
" " 1–4″ " "	35	35	35		—		
4″ " 1–4″ " "	>50				—		
" " 3–4″ " "	—				60		
6″ " 1–4″ " "		>55	55				
3″ " 1–3″ Mid-span Insulator (Spacer)	—	55	55	—		50	
4″ Spacing, 1–4″ Mid-span Insulator (Spacer)	>55	>55	60	55		55	

	Temp. °F. 20	30	40	50	60	70
	Sag— Inches 10.5	11.5	12.5	14.0	15.0	16.0
6″ Spacing, 1–4″ Diam. Disc			45			

260-Foot Span [1]

	Temp. °F. 20	30	40	50	60	70	80
	Sag— Inches 30.0	32.0	34.5	37.5	40.0	43.0	46.0
3″ Spacing, Unequipped Wires	25	25	20	20			
4″ " " "	—	30	25	—			
6″ " " "	45	—	35	—			
3″ Spacing, 2½″ Mid-span Insulator (Spacer)	30	30	35	35		25	25
3″ Spacing, 3″ Mid-span Insulator (Spacer)	—	35	—	35		30	—
4″ Spacing, 3½″ Mid-span Insulator (Spacer)	40	40	40	35		—	
4″ Spacing, 4″ Mid-span Insulator (Spacer)	—	40	—	40		35	

	Temp. °F. 20	30	40	50	60	70
	Sag— Inches 30.0	32.0	34.5	37.5	40.0	43.0
4″ Spacing, Unequipped Wires	30		—		25	
6″ " " "	40		40		35	
8″ " " "	60		—		50	
12″ " " "	>65		—		>55	

NOTES.—1. The number of spans used in obtaining these data were as follows: 100-foot span—one span, 130-foot span—six spans, 160-foot span—one span, 260-foot span—one and two spans.

2. All sag figures given in this table conform to Bell System Practices G31.115.
EQUAL SAG— Both wires of each pair were maintained at the sag given in these Bell System Practices.
UNEQUAL SAG— The wire away from the prevailing winds was sagged in accordance with these Bell System Practices while the windward wire had a greater sag by an amount based on a figure of 3 inches at 60° F. for the 100, the 130, and the 160-foot spans and 6 inches at 60° F. for the 260-foot spans.

3. During the early stages of the wire spacing study no appreciable difference was noticed in the contacting tendencies of 0.104 and 0.165 inch diameter hard drawn copper wires and for this reason it was agreed to use 0.165 inch diameter wire in all subsequent tests. All the data given in this table were obtained with 0.165 inch diameter hard drawn copper.

4. The sign " > " indicates that a velocity as high as the figure given has occurred when the wires were at the sag designated without contacts taking place.

5. Mid-span insulator—A rod type insulator of glazed Isolantite secured to the two wires of a pair at mid-span definitely fixing the wire spacing at that point.

6. Disc—An insulating disc spacer of phenol fabric 3 or 4 inches in diameter secured to one wire of a pair, the wire passing through the center and perpendicular to the disc. The number of discs per span is indicated. When one was used it was located at mid-span. When two were used they were both located on the same wire one at one-third span from each cross-arm. The three-disc arrangement comprised the two disc arrangement with a third disc located at mid-span on the other wire of the pair.

7. Unequipped Wires— Wires not equipped with discs or mid-span insulators.

* Average of instantaneous velocities recorded by a Burton 4-cup anemometer and recorder. The instantaneous velocity registered by this instrument is approximately equivalent to 1.4 times the 5-minute average velocity.

** To obtain some information on the effect of the length of line a few of these wire arrangements were installed on an 18-mile line between Mt. Pocono and Scranton, Pa. The results were in general agreement with those given here.

were obtained when the sags in the wires were small. Threshold velocity data for the cases of larger sags were obtained by repeating the test in some instances with the wires so tensioned that in cold weather the sags were equivalent to those ordinarily prevailing in warm weather.

In considering the threshold velocities given in this article and in Table I the relative frequency of occurrence of these velocities has an important bearing upon the amount of contacting to be expected. In Fig. 9 the curve for maximum (comparable to threshold) wind velocities experienced at Chester, shows that for velocities greater than about 20 miles per hour the frequency of occurrence decreases as the velocity is increased. For example, in terms of five-minute periods, winds in a velocity cell of 36 to 40 miles per hour and those of higher velocity have been found to occur approximately twice as frequently as winds in a cell of 41 to 45 miles per hour and higher. Thus, in the vicinity of Chester, a wire arrangement with a threshold velocity of say 40 miles per hour would be expected to be subjected to winds that would cause contacts during approximately twice as many five-minute intervals in a year as an arrangement with a threshold velocity of 45 miles per hour. At higher wind velocities an increase of five miles per hour in the threshold velocity will be attended by a greater per cent reduction in the number of five-minute intervals during which winds of sufficient velocity to cause contacts will occur.

Since these results were obtained from tests using short lines which were relatively rigid as compared to long lines it was thought that this feature should be given consideration by supplementing these tests with a few representative tests on a longer line. Accordingly, advantage was taken of a toll line, in the Pocono Mountains of eastern Pennsylvania, which was to be dismantled and an 18-mile section of the line was equipped with four pairs of wires each with a different spacing. The pairs were connected to a recorder which registered a contact of practically the same definition as the recorder at Chester. Data from this line were recorded for approximately two winter seasons. The results were in substantial agreement with those obtained from the tests at Chester.

Equilibrium Position of a Span of Wire in Natural Winds

With regard to the theory [1] relative to the equilibrium position of a span of wire under the influence of a steady wind, a study was conducted at the Chester site to investigate the applicability of this theory under the varying conditions of natural winds. Owing to the

[1] Loc. cit.

gustiness of natural winds it was apparent that the study would have to be conducted on a statistical basis. The equilibrium position of a span of wire in a steady wind can be defined by the angle between the vertical plane through the supports and the plane of the suspended wire. This angle is given by equation (1) in Appendix I. Briefly, the problem was to determine this angle for a large number of cases over the complete range of natural wind velocities experienced and to determine the degree of agreement between these values and the angle given by the theory for the corresponding steady wind velocities.

A pair of 0.165-inch diameter hard drawn copper wires with a lateral spacing of 16 inches was installed in a 260-foot span with the supports at the same level. The two wires were maintained at equal sags throughout this study. To prevent movements of the pole supports four guys were used on each pole, three 120° apart attached at about two-thirds the height above the ground and a head guy attached at approximately the top of the pole. The wires were approximately normal to the prevailing northwest winds and were located in close proximity to and at approximately the same height as the graphic recording anemometer and wind vane used in the wire spacing study. A Pathé motion picture camera was modified to take a continuous picture of a point (center of span) on each wire. The camera was mounted rigidly directly under the wires at the center of the span. A fine platinum wire was attached to the camera just above the film to provide a fixed zero reference point on the film and a mechanism was provided for synchronizing the wind velocity chart and the motion picture film.

With this equipment the wires were photographed when at rest, with no wind present, and a number of pictures were obtained at various wind velocities ranging from zero to approximately sixty miles per hour. Figure 10 is a photograph of a section of film showing the wire images and the reference line. Examination of the films disclosed that except on rare occasions when the wind velocity was low the wires were continually in motion and the point photographed on each wire was represented by a wavy line.

In most cases, during the occasions when the wires were being photographed, the wind was approximately normal to the line. The variation in the distance of each wire image on the film from the zero reference line (with reference to the distance when the wires were at rest) provided a means of determining the actual horizontal displacement of the wires for any recorded wind velocity through the use of the ratio of the actual spacing between the wires to the spacing between the wire images on the film, equation (2), Appendix I. The angle of

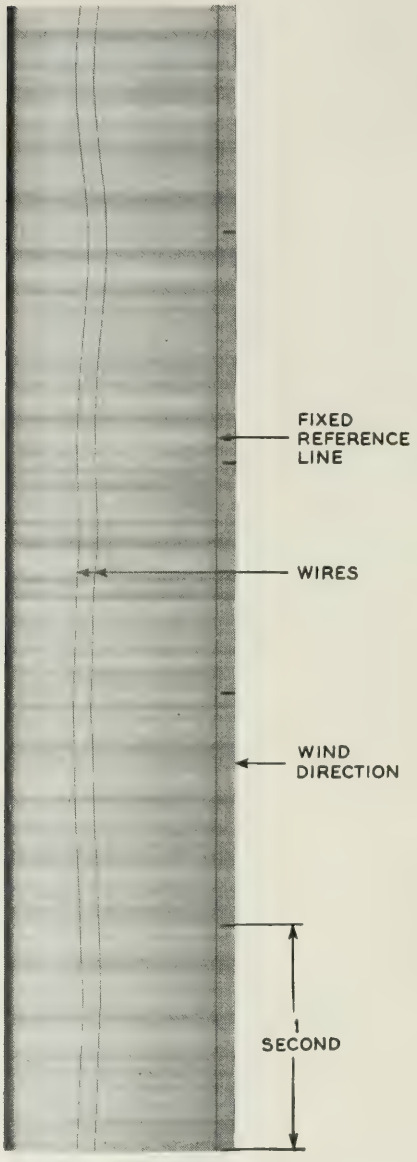


Fig. 10—A photograph of a section of motion picture film showing the transverse movement of wires in wind.

TABLE II

Deflections of Wires in Winds

Transverse Wind Velocity*—M.P.H.	Values of α' Computed from Measurements Made on Motion Picture Films—Degrees																		
	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	50	52	54
Theoretical Angle, α—Degrees	8.0	10.0	12.0	14.0	16.5	19.0	21.5	24.0	27.0	30.0	33.0	36.0	39.0	41.5	44.5	47.0	49.5	51.5	54.0
	11.5	10.0	11.0	10.0	22.0	20.0	17.5	27.0	23.0	29.0	34.5	47.0	47.5	52.5	52.5	50.0	46.0	38.0	45.0
	12.5	9.5	11.0	9.0	21.0	17.0	21.0	28.0	27.5	46.0	39.5	46.5	51.0	45.5	57.5	46.5	41.5	40.0	41.5
	12.0	8.5	10.0	10.0	19.0	19.0	20.0	26.0	22.0	46.0	47.0	46.5	38.0	44.5	59.0	48.5	41.5	41.5	50.0
	8.5	8.0	10.5	10.5	19.0	21.5	17.5	21.5	22.0	29.5	27.0	44.5	37.0	46.5	41.5	44.5	38.5	33.5	58.0
	8.5	9.0	10.5	17.0	19.5	21.5	26.0	24.0	27.5	32.0	29.5	40.0	38.0	46.5	46.5	35.5	40.5	51.0	59.0
	8.0	7.5	10.5	16.5	11.0	20.5	23.5	29.0	25.5	31.0	29.5	40.0	40.5	48.0	41.5	49.5	42.0	48.5	59.0
	7.0	7.5	9.5	16.5	17.5	10.0	21.5	27.5	24.5	31.0	27.0	46.5	39.0	47.5	45.0	47.5	43.0	58.0	49.5
	11.0	9.0	8.5	17.5	17.5	10.0	22.5	28.5	26.5	29.5	26.5	32.5	37.0	45.0	34.0	45.0	53.0	54.0	40.5
	8.0	9.0	7.5	19.0	17.0	22.0	22.0	22.0	26.0	34.5	26.5	36.0	58.5	41.0	35.0	36.5	46.5	48.5	50.0
	8.5	10.0	12.5	19.0	17.5	19.0	23.0	22.5	29.0	30.0	27.5	37.5	48.0	42.5	41.5	35.5	47.0	46.0	60.5
	7.5	9.5	12.0	18.0	17.0	27.0	20.5	19.5	25.5	31.5	26.5	34.5	33.5	45.0	49.5	51.0	47.5	42.0	51.5
	8.5	9.0	12.0	22.5	16.5	27.0	20.0	21.5	25.5	26.5	36.5	33.5	33.5	45.0	44.5	43.5	46.0	50.5	49.0
	8.5	9.5	10.5	19.5	23.0	27.5	20.5	19.5	23.5	25.0	33.0	42.5	26.5	44.0	45.0	33.0	54.0	42.5	48.5
	7.5	8.0	9.5	18.0	24.0	27.0	26.0	28.5	23.0	25.5	35.5	42.0	27.5	38.0	26.5	27.0	54.5	55.5	55.0
	8.0	8.0	8.5	19.0	22.5	13.5	24.0	26.5	21.0	23.5	33.0	40.5	30.0	38.0	63.5	42.0	56.0	47.5	50.0
	8.0	11.0	8.5	17.0	26.0	13.5	21.0	25.5	21.0	22.5	26.5	37.5	32.0	32.0	67.0	48.5	49.0	49.5	49.0
	7.5	9.0	10.0	17.0	15.0	12.5	23.5	30.0	20.0	23.5	36.0	34.5	29.0	25.0	55.5	61.0	65.0	48.0	52.5
	8.0	9.0	10.0	17.5	13.5	13.5	26.5	26.0	29.0	24.0	33.5	34.0	31.5	28.5	44.0	53.0	54.0	56.0	46.5
	7.0	9.0	9.5	16.5	13.5	10.5	25.5	23.5	21.5	24.5	37.0	37.0	46.0	32.0	52.0	48.0	37.0	59.5	54.5
	8.0	8.0	9.0	16.5	13.5	11.0	22.5	21.0	31.0	22.5	38.5	36.5	51.5	34.5	43.5	47.0	36.0	61.0	49.5
Average Value, $\bar{\alpha}'$	8.7	8.9	10.0	16.3	18.3	18.2	22.2	24.9	24.7	29.4	32.5	39.5	38.8	41.1	47.2	44.6	47.0	48.5	51.0
Standard Deviation, σ'	1.695	0.922	1.332	3.705	4.060	6.283	2.674	3.403	3.169	6.892	5.801	4.976	9.217	7.491	10.325	8.086	7.600	7.661	5.659

* Each wind velocity tabulated represents a 2-mile cell, that is, 20 m.p.h. includes all transverse velocities from 19.1 to 21.0 m.p.h. inclusive.

deflection of the suspended wires from a vertical plane in a natural wind could then be computed, equation (4), Appendix I, since the sine of this angle is equal to the horizontal displacement divided by the stretched sag of the wires. If the supports were assumed to be rigid the stretched sag could be calculated directly through the use of equation (3) in Appendix I. However, despite the precautions taken to prevent movements of the pole supports it was found that the poles would bend when the tension in the wires varied. For this reason the deflection of each pole for various wire tensions was measured and this factor was taken into account in determining the stretched sag. The correction applied to the stretched sag as computed from equation (3) was in some cases as great as three inches. Furthermore since the length of wire varies with temperature the particular temperature at which a picture was taken had to be given consideration in computing the stretched sag.

Following the procedure described above, 20 values of the angle representing the equilibrium position of the wires were obtained for each two-mile-per-hour cell of transverse or normal wind velocity over a range of 17 to 55 miles per hour. The average experimental angle was computed for each velocity cell and the degree of dispersion of the individual values was determined in the regular manner.

Table II gives the values of the angle of deflection of the wires calculated from the experimental data, also the average angle and the best estimate of the true standard deviation for each two-mile-per-hour wind velocity cell. For comparison the theoretical angle of deflection as computed through the use of equation (1) is given for each cell. The maximum and the minimum angles determined experimentally might be plotted versus the theoretical angles, but these data furnish no definite measure of the dispersion since maximum and minimum values depend upon the number of observations made. For this reason the degree of dispersion was determined for single observations and for averages by obtaining an estimate of the true standard deviation which is independent of the number of observations. Figure 11 shows the frequency distribution of the angles determined experimentally and also the "three-sigma" limits for the wind velocity cell of 25.1 to 27.0 miles per hour.

In Fig. 12, the average experimental angle for each velocity cell was plotted against the theoretical angle as given by equation (1) and a regression or trend line was determined for these points. For comparison with this line is given a reference line of exact agreement. The "three-sigma" limits for single observations and for averages of 20 observations, were also plotted against the theoretical angle in

Fig. 12. The standard deviation shows a tendency to increase as the angle of deflection is increased. However, as might be expected from the use of natural winds for the experiments in place of the steady wind

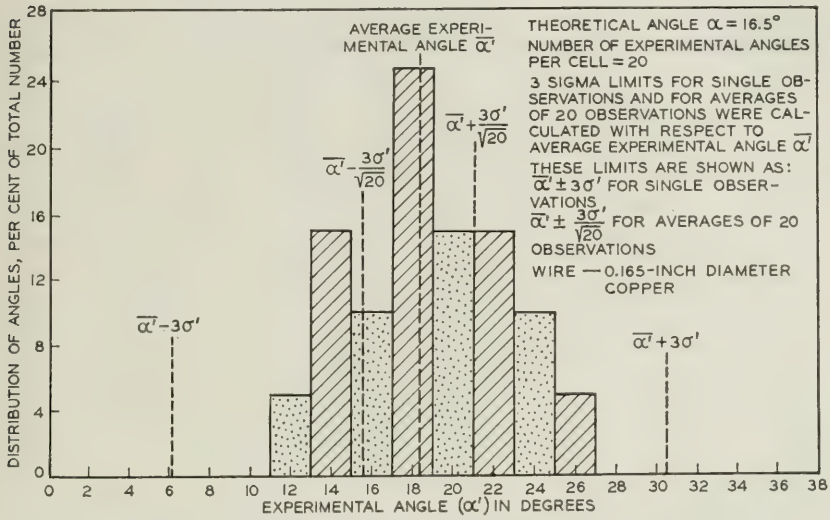

DISTRIBUTION OF ANGLES, PER CENT OF TOTAL NUMBER
EXPERIMENTAL ANGLE (α') IN DEGREES
AVERAGE EXPERIMENTAL ANGLE $\overline{\alpha'}$
$\overline{\alpha'}+\frac{3\sigma'}{\sqrt{20}}$
$\overline{\alpha'}-\frac{3\sigma'}{\sqrt{20}}$
$\overline{\alpha'}-3\sigma'$
$\overline{\alpha'}+3\sigma'$
THEORETICAL ANGLE α = 16.5°
NUMBER OF EXPERIMENTAL ANGLES PER CELL = 20
3 SIGMA LIMITS FOR SINGLE OBSERVATIONS AND FOR AVERAGES OF 20 OBSERVATIONS WERE CALCULATED WITH RESPECT TO AVERAGE EXPERIMENTAL ANGLE $\overline{\alpha'}$
THESE LIMITS ARE SHOWN AS: $\overline{\alpha'} \pm 3\sigma'$ FOR SINGLE OBSERVATIONS, $\overline{\alpha'} \pm \frac{3\sigma'}{\sqrt{20}}$ FOR AVERAGES OF 20 OBSERVATIONS
WIRE — 0.165-INCH DIAMETER COPPER

Fig. 11—Distribution of experimental angles of deflection of a suspended wire in a transverse wind velocity cell of 25.1 to 27.0 miles per hour.

assumed in the theory, there was irregularity in the "three-sigma" limits. For this reason a regression line was determined for all the points in each particular "three-sigma" limiting group and the lines were drawn. This graph shows the agreement between the angle of deflection of the wire as determined by the experimental method and that given by the theoretical equation (1). The area between the two extreme limit lines represents approximately the range within which single values determined experimentally would be expected to fall. Likewise, the area between the two inner limit lines represents approximately the range within which averages of 20 values for a particular wind velocity cell would be expected to fall.

In general, the results indicate that the theory for the equilibrium position of a suspended wire under the influence of an assumed steady transverse wind is applicable within reasonable limits to a wire subjected to the varying conditions of natural winds.

An Accelerated Method of Test

Testing the merits of various arrangements in natural winds is a slow procedure in which it is necessary to await the course of nature.

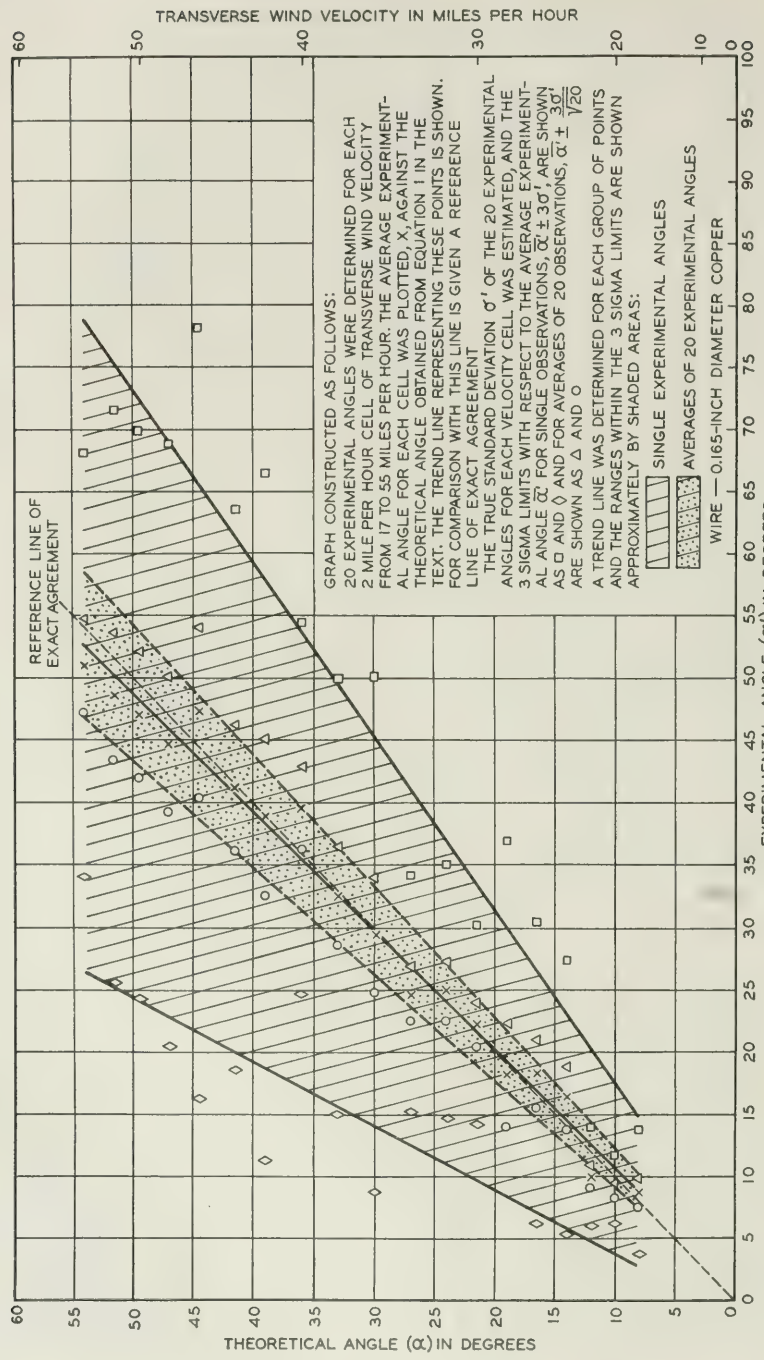

TRANSVERSE WIND VELOCITY IN MILES PER HOUR
REFERENCE LINE OF EXACT AGREEMENT
GRAPH CONSTRUCTED AS FOLLOWS:
20 EXPERIMENTAL ANGLES WERE DETERMINED FOR EACH 2 MILE PER HOUR CELL OF TRANSVERSE WIND VELOCITY FROM 17 TO 55 MILES PER HOUR. THE AVERAGE EXPERIMENTAL ANGLE FOR EACH CELL WAS PLOTTED, X, AGAINST THE THEORETICAL ANGLE OBTAINED FROM EQUATION 1 IN THE TEXT. THE TREND LINE REPRESENTING THESE POINTS IS SHOWN. FOR COMPARISON WITH THIS LINE IS GIVEN A REFERENCE LINE OF EXACT AGREEMENT
THE TRUE STANDARD DEVIATION σ' OF THE 20 EXPERIMENTAL ANGLES FOR EACH VELOCITY CELL WAS ESTIMATED, AND THE 3 SIGMA LIMITS WITH RESPECT TO THE AVERAGE EXPERIMENTAL ANGLE ᾱ' FOR SINGLE OBSERVATIONS, ᾱ' ± 3σ', ARE SHOWN AS □ AND ◇ AND FOR AVERAGES OF 20 OBSERVATIONS, α' ± 3σ'/√20 ARE SHOWN AS Δ AND ○
A TREND LINE WAS DETERMINED FOR EACH GROUP OF POINTS AND THE RANGES WITHIN THE 3 SIGMA LIMITS ARE SHOWN APPROXIMATELY BY SHADED AREAS:
SINGLE EXPERIMENTAL ANGLES
AVERAGES OF 20 EXPERIMENTAL ANGLES
WIRE — 0.165-INCH DIAMETER COPPER
THEORETICAL ANGLE (α) IN DEGREES
EXPERIMENTAL ANGLE (α') IN DEGREES

Fig. 12—A comparison of experimental and theoretical angles of deflection of a suspended wire in winds.

In order to narrow the scope of the test in natural winds to the most representative and promising of the various proposed arrangements, an accelerated method of test was used for characterizing the proposals in a preliminary way.

This accelerated method was suggested by the above mentioned theory concerning the equilibrium position of a span of wire in wind. From this theory it was conceived that the relative merits of a type of wire arrangement could be determined by successively deflecting one wire of a pair outward and upward and releasing it to swing towards the other wire through an increasing series of known angles until the two contacted. The theoretical wind velocity corresponding to this minimum angular displacement producing contacts would then be determinable through the use of equation (1).

While it was realized that such a procedure does not simulate the contacting of wires in natural winds, it was thought that from a relative point of view the arrangement requiring the greater angular displacement and therefore the higher theoretical velocity to produce contacting might also require a higher natural wind velocity. Further, it was felt that there was a possibility of being able to determine not only the relative merits of types of arrangements but also their natural wind threshold velocities by correlating the data obtained by the accelerated method with those obtained in the natural wind tests.

Accordingly, a number of tests were made using arrangements on which considerable natural wind data were available. The test set-up comprised a pair of 0.165-inch diameter hard drawn copper wires installed in proximity to the ground with suitable means for changing spacings and sags at will. In order that the deflection of the wire might be accurately controlled and the results reproducible, a series of rods used as guides were mounted rigidly on a vertical support at the center of the span in a plane at right angles to and in proximity to the wires. The points of these rods were positioned in the arc of a circle with a radius equal to the sag in the manually deflected wire. The intervals between the points of the rods in terms of theoretical wind velocity were five miles per hour. The arrangement of rods represented a range of 20 to 80 miles per hour. The test apparatus is shown in the accompanying Fig. 13.

In the first stages of these tests it was found that a pair of wires would not always contact for the same minimum angle of deflection. Experiments showed, however, that during the absence of any natural wind there was a minimum angle for a given wire arrangement which would produce contacting five times in five consecutive trials. This refinement of the method was then adopted and the theoretical

velocity equivalent to this angle was termed the accelerated method threshold velocity or the velocity at which contacting begins. It is, therefore, not necessarily the minimum angle or velocity at which the wires can be made to contact but the minimum angle or velocity at which the wires will contact for each of five consecutive trials in practically still air.

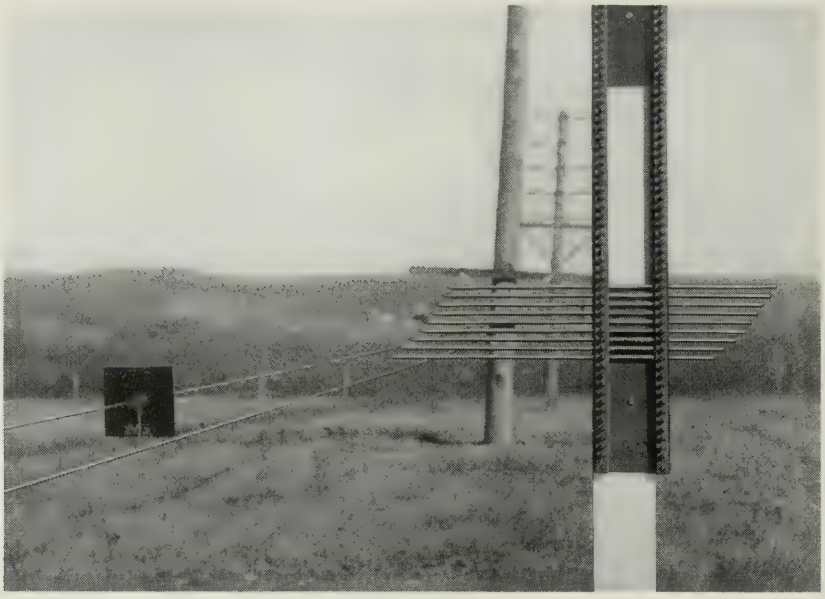


Fig. 13—Apparatus for the accelerated test.

In order that the contacts recorded by the accelerated method should be comparable in definition to those recorded by the natural wind method, the wires used in the accelerated test were connected to the same recording apparatus as was used in the natural wind tests.

From the results obtained through the use of the accelerated method of test an empirical equation was developed for the case of a pair of wires with equal sags. This equation (6) is given in Appendix II. The comparable empirical equation (5) for natural winds, referred to above, is also given in Appendix II. With these two equations it is possible to determine the expected natural wind threshold velocities of a wire arrangement through the use of the accelerated method of test. An equation (7) for this use, which was obtained from the above two equations (5) and (6), is also given in Appendix II.

While empirical equations were developed only for the case of a

pair of wires with equal sags, the applicability of the accelerated method is not limited to this case. It was also used where the sags in the two wires of a pair were unequal and, as referred to later, for determining the most promising design of insulating disc (described below) for use in natural wind tests as a means of mitigating the contacting on pairs with the wires spaced, 3, 4 and 6 inches.

Anti-Contacting Insulators

As stated above it was found that wires spaced 3 or 4 inches contacted in wind velocities rather commonly experienced. Some

Fig. 14—Insulating disc.

contacting was also recorded on pairs with wires spaced 6 inches. In giving consideration to means of increasing the natural wind threshold velocities of such circuits to those occurring less frequently, two types of anti-contacting insulators were developed. One type, Fig. 14, was a perforated disc of insulating material. When this type was installed on one wire of a pair it was not in contact with the other wire of the pair except when forced there by the action of the wires in wind. The other type, Fig. 15, was a rod-shaped insulating spacer. This spacer bridged the two wires of a pair in the span.

The insulating discs used were 3 and 4 inches in diameter. The arrangements of these discs tested in natural winds comprised one, two or three discs per span per pair of wires. When one disc was used, it was placed at the approximate center of the span on the wire

of the pair to the windward side of the line. When two discs were used they were placed on the windward wire at one-third of the distance from each support. The three-disc arrangement comprised the two-disc arrangement with the third disc placed at the center of the span on the other wire of the pair.

Fig. 15—Insulating spacer.

In selecting the disc and these sizes, various shapes and sizes of insulators (one of which is shown in Fig. 13) were tested using the accelerated method of test referred to above. The circular type was found to give as good results as any other shape and had the advantage of being simple in design.

The insulating spacers were used one per span per pair of wires located at the approximate center of the span. Any insulating spacer bridging the wires of a pair constitutes an additional line leakage path and from this standpoint is undesirable. In this respect the discs have a distinct advantage over the spacers. As stated above, they are not normally in contact with both wires and when such contacts take place they are generally of short duration. Then too, in a long line with wires equipped with discs it is improbable that more than a short section would be affected at one time. The thought was that owing to the additional line leakage provided by the spacers, their use would be confined to the occasional long span and the use of discs to the shorter spans. For this reason the tests involving discs were confined to spans of 160 feet and less and those involving spacers to spans of 160 and 260 feet. The dimensional characteristics

of the spacers important from the standpoint of mitigating contacting were tested. These related to the length of the spacer between wires. It was not known whether this distance should be equal to, less than or greater than the spacing between the wires of a pair at the crossarm. The first tests indicated that when this distance was ½ inch greater than the wire spacing, there was a tendency for the pair to roll and the wires to twist around each other. Later tests were confined, therefore, to spacers with a distance between the wires the same as the spacing or ½ inch less. The data obtained in the natural wind tests on wire arrangements where discs and spacers were used are given in Table I.

The effect of insulating discs was to increase the threshold velocities over those for similar arrangements without discs by about 5 to 20 miles per hour. The 4-inch diameter disc has some advantage over the 3-inch diameter disc. Three discs per span or even two discs give results somewhat better than those obtained with a single disc but the gain is relatively slight.

The spacer which holds the wires in the center of the span the same distance apart as the spacing at the crossarm, in general, increased the threshold velocities about 10 to 30 miles per hour for pairs with wire spacings of 3 and 4 inches at equal sags and suspended in spans of 160 and 260 feet over comparable arrangements of unequipped wires. No information is available for the case where the wires of a pair have unequal sags.

The data for the 160-foot span give some comparison between the effectiveness of discs and spacers. In the case of wires spaced 3 inches with one 4-inch diameter disc, contacts occurred at a threshold velocity of 35 miles per hour while the comparable figure for the spacer was 55 miles per hour.

Abstract Conclusion

In these tests typical toll telephone wires were placed in spans of 100, 130, 160 or 260 feet with sags of 4 to 45 inches (depending upon the span length and temperature) and with horizontal spacings between the wires of a pair of 3, 4, 6, 8 or 12 inches. It was found that during the absence of glaze the wind velocities normal to the line when swinging contacts began to occur increased with the wire spacings and also with span lengths (if wire tensions were increased so as to maintain a given sag) and decreased when the sag was increased. An empirical equation based upon this relation and the data (Table I) has been developed for the case of wires at equal sags. As a brief example of the results, in the case of a 130-foot span, it was found that wires

spaced 12 or 8 inches were practically free from swinging contacts in wind velocities below about 70 miles per hour, wires spaced 6 inches contacted at velocities around 50 miles per hour and wires spaced 4 or 3 inches contacted at the more common velocities of 30 or 40 miles per hour.

In the absence of glaze, when the wires of a pair were at unequal sags, about 3 inches difference in the 130-foot spans and about 6 inches in the 260-foot spans, there was in general a somewhat greater tendency toward contacting in the shorter spans and a lesser tendency in the longer ones than when the wires were at equal sags.

When glaze was on the wires their action was more erratic and swinging contacts were more general and occurred at lower wind velocities than when glaze was not present.

Regarding the theory [1] relating to the equilibrium position of a suspended wire in a steady wind, tests were conducted at Chester, New Jersey, in which the displacements of a copper wire were photographed in natural winds. It was found that there was general agreement between the angle of deflection of a wire as determined by this theory for a given steady wind and the angle obtained experimentally in a comparable natural wind.

The experimental substantiation of this relationship led to the development of an accelerated method for a quick and economical preliminary classification of various wire arrangements. This method was useful in selecting from among several similar arrangements the most promising ones for test in natural winds. An empirical equation based on the data obtained by the accelerated method for the case of equal sags was developed for expressing the relationship between accelerated method wind velocity, span length, wire spacing and sag. By combining this equation with that developed from the natural wind data a third equation was obtained which was used in determining expected natural wind threshold velocities from the accelerated method results.

In regard to the anti-contacting devices included in the study with the wires spaced 6 inches and less, it was found that:

1. A 4-inch diameter insulating disc placed at the approximate center of the span on one wire of the pair increased the normal wind velocities at which contacting began by 5 to 20 miles per hour over those for the same arrangements unequipped. The use of three discs per span or even two gave an improvement over the use of one but the gain was relatively slight.

[1] Loc. cit.

2. A rod-type insulating spacer used to bridge the wires of a pair in the approximate center of the span was somewhat more effective than one disc, increasing the normal wind velocities at which swinging contacts began by about 10 to 30 miles per hour over those for the same arrangements unequipped.

In general, the higher the threshold velocity, the less frequent will be the occurrence of those winds which will cause contacting. Therefore, if the threshold velocity of a wire arrangement is increased 5 or 10 miles per hour or more by the addition of an anti-contacting device there will be a decrease in the amount of contacting occurring dependent upon the original threshold velocity and the amount of the increase. For example, in the vicinity of Chester, New Jersey, an increase in the threshold velocity of a wire arrangement from 40 to 45 miles per hour results in a reduction of about 50 per cent in the number of five-minute intervals during which winds of sufficient velocity to cause contacts will occur. At higher wind velocities an increase of five miles per hour in the threshold velocity will produce a greater per cent reduction in the number of five-minute intervals during which winds of sufficient velocity to cause contacts will occur.

In regard to the field installations with a spacing of 8 inches between the wires of a pair, referred to at the beginning of this article, no serious difficulties have been encountered except in certain sleet areas where there has been some wrapping and freezing together of the wires. In these locations insulating spacers have been installed on a few pairs and their behavior is being followed. The installations in which a 6-inch spacing has been used have been confined to the warmer sections of the country and no serious trouble has yet been encountered.

APPENDIX I

The expression for determining the angle of deflection or equilibrium position of a suspended wire in a steady transverse wind as given in the theory [1] is

$$\tan \alpha = \frac{k V^2}{mg \cos \gamma}, \qquad (1)$$

where α = Angle between the plane of the suspended wire and a vertical plane through the supports,

V = Steady transverse wind velocity (miles per hour),

m = Mass of unit length of wire (slugs),

g = Acceleration of gravity (feet per second per second),

k = Ratio of wind pressure per unit length of wire to square of velocity, and

γ = Angle of inclination of line through supports to the horizontal.

In the study to determine the extent to which this theory would check under the varying conditions of natural winds only the case where $\gamma = 0°$ (both supports at the same level) was considered.

The camera used in this study was equipped with a Tessar F-3.5 lens with a nominal focal length of 40 mm. The films were analyzed with the aid of a motion picture projector. In this projector the film passed over a glass plate on which was engraved a linear scale graduated in hundredths of an inch. The pictures, together with the graduated scale, were projected on to a screen. This method provided a ready means of determining the horizontal displacement of the wire images on the film. The actual wire displacement was then determined through the use of the following relationship:

$$\frac{L_1}{L_2} = \frac{D_1}{D_2}, \tag{2}$$

where L_1 = Distance from wires to camera lens,
L_2 = Distance from camera lens to film,
D_1 = Spacing between wires, and
D_2 = Spacing between wire images on film.

The two wires were maintained at equal sags throughout this study. The equation for determining the stretched sag of a wire if the supports are assumed to be rigid is as follows:

$$a^3 + \frac{3L}{8}(L - R)a = \frac{3wL^4}{64AE}. \tag{3}$$

where a = Stretched sag,
R = Unstressed length of wire,
L = Span length,
A = Cross-sectional area of wire,
E = Modulus of elasticity, and
w = Resultant of wind pressure and gravity components.

As explained in the text, even though the poles were strongly guyed, the supports moved when the tension in the wires varied and corrections were applied to the stretched sag to take account of this movement.

After determining the horizontal displacement and the stretched sag of the wires the experimental angle of deflection (angle of equilibrium position) was calculated by means of the equation:

$$\sin \alpha' = \frac{L_3}{a'}, \tag{4}$$

where L_s = Horizontal displacement of wires,
a' = Corrected stretched sag of wires, and
α' = Angle of deflection determined experimentally as distinguished from the theoretical angle (α).

The details of the method of determining the equilibrium position of the wires for a particular natural wind velocity were as follows:

The horizontal displacement of the two wire images on the film at each wave crest and trough was determined. The displacement for the two wire images at each crest and at each trough was averaged. Next the mean displacement on the film of the crest and trough was calculated and also the mean velocity * was determined for this particular time interval. The mean displacement on the film was then converted to actual wire displacement through the use of equation (2) and the experimental angle was determined by equation (4). This average angle was taken as the equilibrium position of the wires for this mean velocity.

APPENDIX II

Empirical Equations

From natural wind tests on arrangements of wires in which both wires of a pair were maintained at equal sags it was found that in the absence of glaze threshold velocities increase with the spacing between the wires and the span length and decrease as the sag increases. An empirical equation obtained from an analysis of the results is as follows:

$$V_w = 22.4 \left[\frac{L^{0.1} S^{0.3}}{d^{0.25}} \right]^{2.1}, \qquad (5)$$

where V_w = Natural wind threshold velocity (miles per hour),
L = Span length (100 to 260 feet),
S = Wire spacing (3 to 12 inches), and
d = Sag of wires at rest (4 to 45 inches).

The data upon which this equation was based comprised approximately fifty cases where swinging contacts actually occurred. Regarding the degree to which this equation represents these data, there were only about five cases which deviated as much as five miles per hour in terms of threshold velocity and of these only one deviated as much as seven miles per hour. The nomogram given in Fig. 8 was constructed for this equation (5).

* When the direction of the wind was not normal to the line the normal component of the velocity was determined by multiplying the wind velocity by the cosine of the angle between the actual direction of the wind and the normal to the line.[2]

The comparable empirical equation for the accelerated method of test is

$$V_m = \frac{10Y}{1 - 0.692Y}, \tag{6}$$

where

$$Y = \frac{L^{0.05}S^{0.2}}{d^{0.2}},$$

and V_m = Accelerated method threshold velocity.

The other terms are the same as given above. This equation has a degree of accuracy comparable to equation (5).

These two equations, (5) and (6), were combined to form equation (7) which was used to determine expected natural wind threshold velocities from the accelerated method results.

$$V_w = V_m \left[\frac{2.24L^{0.16}S^{0.43}}{d^{0.325}} \right] \left[1 - 0.692 \frac{L^{0.05}S^{0.2}}{d^{0.2}} \right]. \tag{7}$$

The terms in this equation are the same as those given above.

Variable Equalizers

By H. W. BODE

THE use of equalizing structures to compensate for the variation in the phase and attenuation characteristics of transmission lines and other pieces of apparatus is well known in the communication art. Ordinarily, of course, an equalizer has a definite characteristic fixed by the apparatus with which it is to be associated. It may happen, however, that the characteristics demanded of the equalizer cannot be prescribed in advance, either because the characteristics of the associated apparatus are not known with sufficient precision, or because they vary with time. Examples are found in the equalization of transmission lines the exact lengths of which are unknown, or the characteristics of which may be affected by changes in temperature and humidity.

In recent years the problem of providing equalizers which will meet such conditions as these has assumed particular importance because of the large variations in line attenuation which may result from temperature changes in new carrier systems. In some of these the maximum change in attenuation is more than 1 db per mile. Evidently, if a reasonable standard of quality is to be maintained for systems the overall length of which may be several hundred or several thousand miles, these variations must be compensated for with great accuracy. Moreover, since the total amount of correction must necessarily be divided up into much smaller amounts appearing at many points, and since the daily cycle of temperature changes may be large, it is almost essential that the adjustments made be so simple that they can readily be performed automatically by a suitable auxiliary circuit.

The variable equalizers described here are attempts to meet this problem. In order to secure the maximum simplicity it has been assumed that the characteristics of the structure are controlled by a single variable element, which in most cases is a variable resistance. It has also been assumed that the temperature coefficient is known as a function of frequency and that it is the same at all temperatures. What is required, then, is a structure by means of which an arbitrary multiple of a given attenuation characteristic can be introduced into a circuit by changes of a single element.

For purposes of future discussion it is convenient to express this requirement in precise form. If the network functions ideally its loss must be given by an equation of the following type:

$$\theta = F_1(\omega) + F_2(\omega)F_3(R), \tag{1}$$

where R is the variable resistance. The function $F_2(\omega)$ corresponds to the temperature characteristic. It must evidently be under our control. The function $F_1(\omega)$ represents a fixed loss, analogous to that of an ordinary equalizer. It is of less importance since it can always be changed by the addition of separate fixed networks. In several of the structures to be described, however, it also is under our control, so that the networks can be used as combined fixed equalizers and temperature correcting devices. The function $F_3(R)$ expresses merely the calibration of the controlling element with respect to temperature, and its exact form is consequently of minor importance.

It is not difficult to find circuits which function broadly in the manner described by equation (1). In most instances, however, the desired proportionality in the set of variable characteristics is realized only very approximately. A simple circuit, which it is hoped is both a fair and a plausible example, may illustrate the sort of performance to be expected.[1] The structure consists of a condenser in series with a variable resistance, bridged across a resistance circuit, as shown by Fig. 1. For high values of the variable resistance we may anticipate that the attenuation will be low at all frequencies, while at lower values a characteristic rising with frequency should be obtained. The network should then behave much like a radio "tone control." An inspection of the actual characteristics, also shown on Fig. 1, indicates, however, that although this general behavior is in fact obtained, the curves change shape rapidly in every range except that corresponding to high resistances and high frequencies, where they are almost constant.

The distortion exemplified by the curves of Fig. 1 is the greatest obstacle in the design of variable equalizers to satisfy the specifications of equation (1). To a certain extent it is unavoidable. It is easily shown for example, that the transfer admittance from generator to load impedance in any network containing a single variable resistance can be written as

$$Y = \frac{ZY_s + RY_0}{Z + R}, \tag{2}$$

[1] More elaborate circuits have, of course, been devised in the past. Mention should be made in particular of the structure described in U. S. Patent No. 2,019,624, issued to Mr. E. L. Norton. For moderate ranges of variation, the characteristics of this network are somewhat similar to those of the structure exemplified later by Fig. 6. Another method of attack is shown by U. S. Patent No. 2,070,668, issued to Mr. W. R. Lundry.

where R represents this resistance, Z represents the impedance which it faces, and Y_s and Y_0, as the equation implies, are the transfer admittances obtained by short-circuiting and open-circuiting R. It is evident by inspection that log Y, which represents the θ of equation (1), cannot be written in the form which the right-hand side of (1) demands. A certain amount of distortion of the type shown by Fig. 1 must therefore always occur.

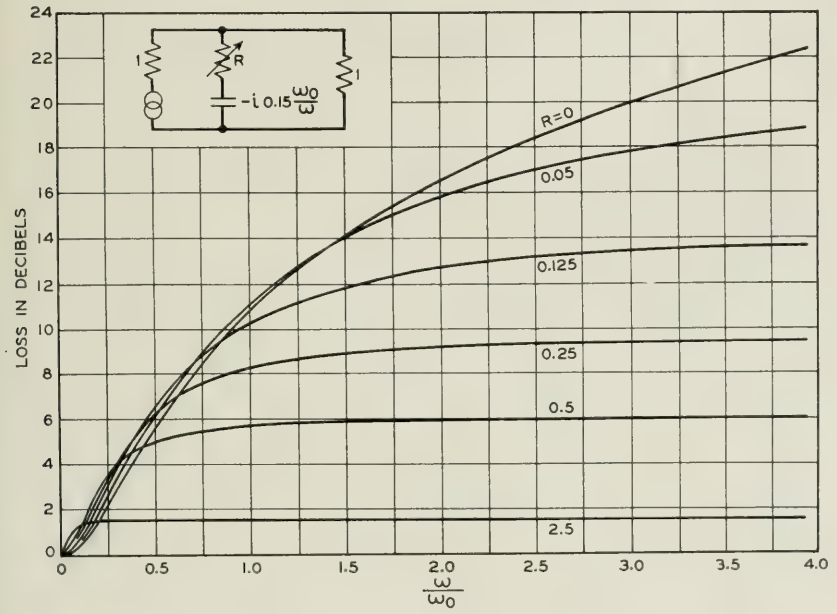


Fig. 1—Characteristics of a simple variable structure.

There still remains the possibility, however, of obtaining a network in which the distortion can be kept within tolerable limits over a given range. The quantities Y_s, Y_0, and Z, which are, of course, all functions of frequency, allow us to determine the transfer admittance at three values of R. The transfer admittances at other settings will then be fixed. If we suppose for simplicity, that the extreme characteristics, corresponding to Y_s and Y_0, are set by the engineering requirements on the structure, the problem reduces to that of so choosing Z in relation to these quantities that the distortion is as small as possible at intermediate settings of R.

Of the variety of possibilities open in the selection of Z, one in particular commends itself by the simplicity and symmetry of the results to which it leads. It is given by the condition

$$Z = \sqrt{\frac{Y_0}{Y_s}} R_0, \tag{3}$$

where R_0 is an arbitrary constant which represents, physically, a reference value for the variable resistance. With the help of this condition (2) becomes

$$Y = \sqrt{Y_0 Y_s} \frac{R_0 + \frac{R}{R_0} Z}{Z + R}, \tag{4}$$

which can be rewritten in a slightly different notation as

$$e^{-\theta} = e^{-\theta_0} \frac{1 + xe^{-\varphi}}{x + e^{-\varphi}}, \tag{5}$$

where $e^{-\theta}$, $e^{-\theta_0}$, x and $e^{-\varphi}$ stand respectively for the quantities Y, $\sqrt{Y_0 Y_s}$, $\frac{R}{R_0}$ and $\frac{Z}{R_0}$.

The significance of the assumption made in equation (3) is apparent from an inspection of equation (5). When $R = R_0$ the total loss θ of the circuit is equal to θ_0. The quantity θ_0 can therefore be described as the average or reference loss of the circuit, corresponding to the average or reference value of R. It is represented by the middle curve shown on Fig. 2. Setting $R = 0$ or $R = \infty$ gives the symmetrically located extreme curves $\theta_0 \pm \varphi$ also shown in the figure. The quantity φ is therefore the extreme change in the attenuation of the network produced by variations in R. Since the two extreme curves correspond to Y_0 and Y_s the situation can also be described by saying that condition (3) fixes the third arbitrary transfer admittance characteristic symmetrically between the first two.

It is easily shown that any other pair of characteristics which correspond to reciprocal values of x, such as those shown by the broken lines in Fig. 2, will also be symmetrically placed with reference to θ_0. This line therefore divides the complete family of characteristics into two equal halves. The departures of the intermediate characteristics from θ_0 are, of course, not strictly proportional to φ. The error can, however, readily be investigated by expanding (5) as a power series in terms of φ.[2] We find

$$\theta = \theta_0 + \frac{x - 1}{x + 1} \varphi + g_3(x)\varphi^3 + g_5(x)\varphi^5 + \cdots, \tag{6}$$

even terms being absent because of the symmetry of the original ex-

[2] A more detailed treatment of this analysis will be found in the writer's U. S. Patent 2,096,027.

pression. The particular forms of the functions $g_3(x)$ and $g_5(x)$ are not of great interest. It is important, however, to know their maximum values, which are respectively 0.03 and 0.002.

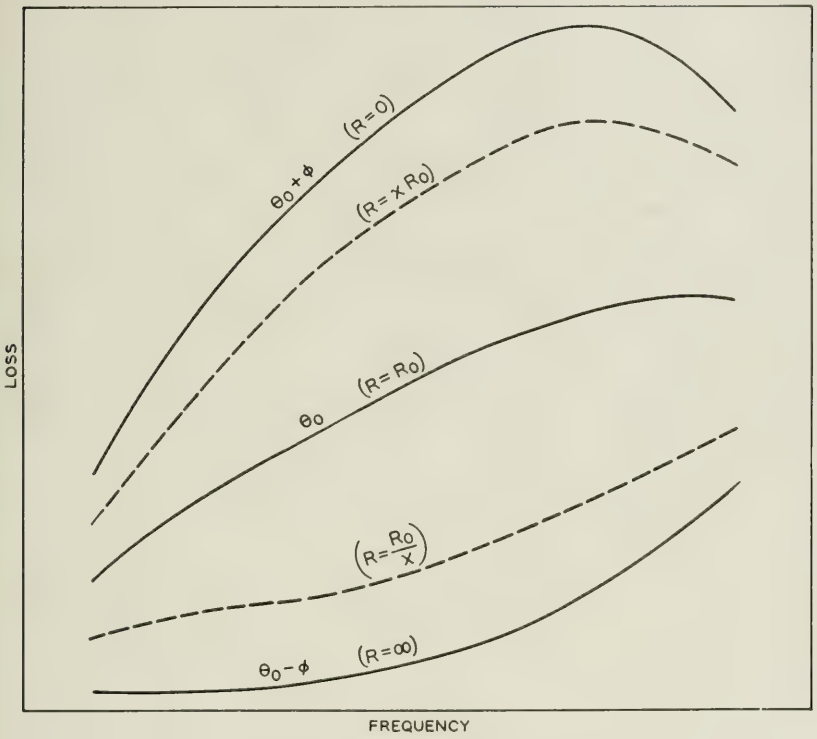


FIG. 2—Diagram to illustrate symmetrical characteristics obtained from the equalizers described in the paper.

The first two terms in equation (6) can evidently be identified with the quantities appearing on the right-hand side of the ideal regulator equation (1). So far as these terms are concerned, the variable characteristics are always strictly proportional to φ. The departure from the ideal is measured by the remaining terms of (6). For ordinary values of φ the series converges so rapidly that only the term in $g_3(x)\varphi^3$ is important. Since the maximum value of $g_3(x)$ is known an estimate of the distortion is easily made, although some allowance is necessary for the possibilities of "splitting differences" in the design process. With the help of these allowances, however, it turns out that the distortion should be about 0.1 db when the maximum value of φ is 1 neper, corresponding to a total variation in attenuation of about

18 db. For other values of φ the distortion is, of course, proportional to the cube of the total change in attenuation.

This estimate is in good agreement with the computations made on actual networks. It also covers the range of greatest practical interest, since in most communication systems changes in attenuation of as much as 18 db produce undesirable variations in the levels of the different channels with respect to one another or to interfering signals. If necessary, however, it appears to be possible to go considerably farther. This possibility arises from the fact that in actual practice φ will be complex, which means that the structure acts as a variable equalizer with respect to both phase and attenuation. Ordinarily, however, only the variation of the attenuation characteristic is of interest. Since the real component of φ^3 depends upon both the real and imaginary components of φ, it is thus possible, by choosing the proper relation between these latter two quantities, to eliminate, effectively, the third order term as well as the even order terms in the general expansion. The first disturbing term is then of the fifth order, and has a very small coefficient. If we assume that the desired relation between the real and imaginary components of φ can be obtained with sufficient precision in a physical network, it appears that this process allows us to confine the distortion to 0.1 db for total variations in attenuation as great as 30 or 35 db. Since the distortion now depends upon the fifth power of the total variation in attenuation it is, of course, very small for more moderate variations. For example, under the same assumptions it is only about 0.001 db for a total variation of 12 db.

All of these relations, of course, have no utility unless structures meeting the general conditions laid down by equation (3) can be found. The simplest structure for the purpose appears to be the Π of fixed resistances shown in Fig. 3. A set of illustrative characteristics, drawn on the assumption that the parameter a equals 2, is shown at the bottom of the figure.

At first sight, this may appear to be a trivial illustration, since θ_0 and φ are merely constants, and the structure thus has only the properties of an ordinary gain control. It is possible, however, to introduce auxiliary networks by means of which θ_0 and φ can be made prescribed functions of frequency. For example, θ_0 can be altered by adding an ordinary equalizer in tandem with either terminating resistance. The modification which allows us to vary φ may be somewhat less obvious. It consists of the introduction of a symmetrical four-terminal network having the image impedance R_0, between the variable resistance and the terminals to which it was previously connected, as shown by Fig. 4.

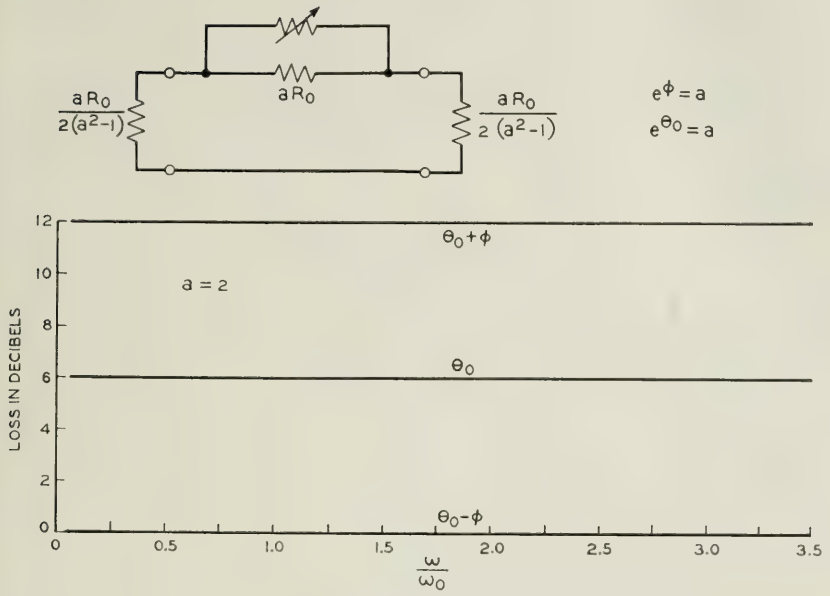


Fig. 3—The simplest type of symmetrical variable equalizer.

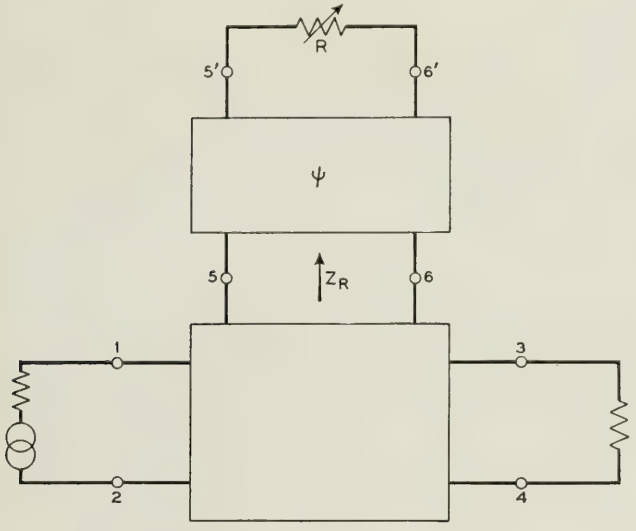


Fig. 4—Adjustment of the variable characteristic by the addition of an auxiliary network.

The effect of the added network is easily understood from the preceding equations. It will be noticed that although these equations were written under the assumption that R is a real quantity, they will still be valid if R is complex. We need therefore merely to replace R by the impedance of the auxiliary network terminated by the variable resistance. If we represent this impedance by Z_R, the appropriate expression is

$$Z_R = R_0 \frac{x + \tanh \psi}{1 + x \tanh \psi}, \tag{7}$$

where ψ is the transfer constant of the added network and x is, as before, the ratio of the variable resistance to R_0. Since reciprocal values of x still correspond to reciprocal values of $\frac{Z_R}{R_0}$, all of the preceding conditions of symmetry in the resulting family of characteristics are maintained. The simplest formulation for the new φ is secured from equation (6). Upon replacing the "x" of this expression by $\frac{Z_R}{R_0}$ we readily find that the equation becomes

$$\theta = \theta_0 + \frac{x - 1}{x + 1} e^{-2\psi} \varphi + \text{higher order terms.} \tag{8}$$

The effect of the added network is therefore merely to multiply the original φ by $e^{-2\psi}$.

An example of the use of this device in conjunction with the network of Fig. 3 is given by Fig. 5. The parameter a was chosen equal to 2, which corresponds to a maximum change in attenuation of 12 db. The auxiliary network, as will be seen, is a conventional bridged-T equalizer. The characteristics of the structure are shown by Fig. 6. The series of straight lines represents the assumed curves, while the circles show the actual computed points. The scale of the drawing is too small to show the differences between the two very clearly. The actual error, at the worst setting and frequency, amounts, however, to about 0.05 db. Of this total, about half is due to the intrinsic distortion of the structure, which in this instance is controlled by the φ^3 term, the effect of the imaginary component of φ being negligible. The remaining half is due to the failure of the bridged-T network to realize the desired ψ characteristic with sufficient precision, and could presumably be eliminated by the addition of more elements to the structure.

Since both θ_0 and ψ can be controlled by auxiliary networks, the structure of Fig. 3 is by itself theoretically sufficient to meet all requirements. There exist, however, a number of other circuits which also

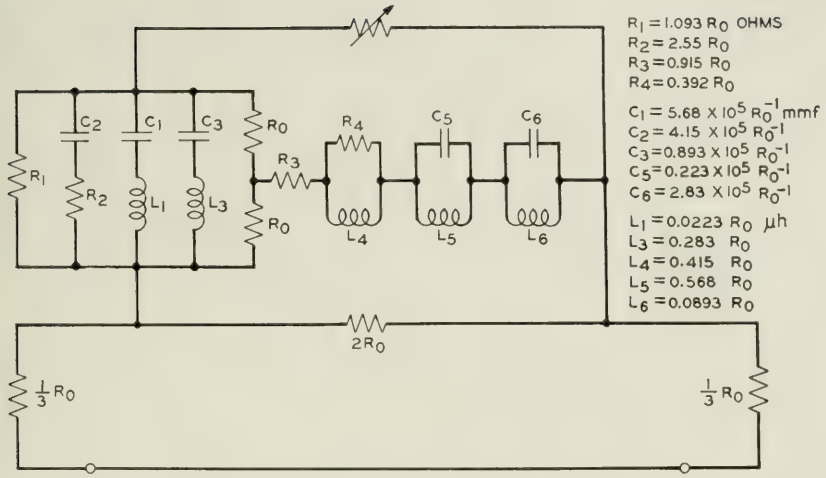


Fig. 5—An equalizer of the type shown in Fig. 3 after the addition of an auxiliary network.

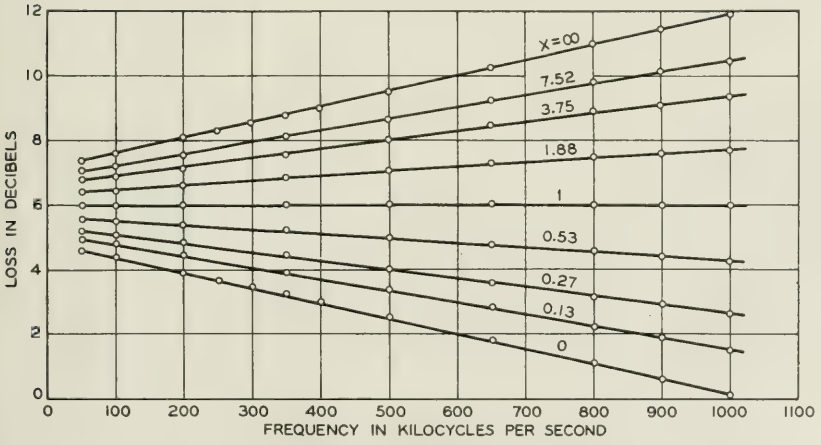


Fig. 6—Characteristics of the structure shown by Fig. 5.

satisfy the condition expressed by equation (3). For example, any structure having the general configuration of Fig. 7 will meet this condition provided the impedances Z_1, Z_2 and Z_3 are so related that when the network is considered as a 4-terminal structure transmitting from a–a' to b–b' it has a constant resistance image impedance equal to R_0 at a–a'.

In contrast to the network of Fig. 3 in which both θ_0 and φ are merely constants, most of the networks which have been found give rather complicated expressions for these two quantities. There are still a number, however, the properties of which are sufficiently simple to be of special interest. The first two are shown by Fig. 8. They have

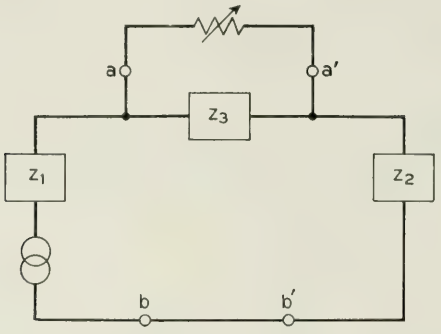


Fig. 7—Diagram to illustrate requirements on a more general form of variable equalizer.

the same design formulae, but one is terminated in a finite resistance at both ends, while an open-circuit, such as the grid of a vacuum tube, must be provided at one end of the other. Illustrative characteristics, drawn on the assumption that the impedance Z_{11} is a simple inductance are shown at the bottom of Fig. 8. It will be seen that φ is still a constant, so that an additional network must be added to control it, exactly as in the structure of Fig. 3. The reference loss θ_0, however, now varies with frequency and can be controlled by the adjustment of Z_{11}. The design impedances have been written as Z_{11} and Z_{21} in accordance with the usual convention for fixed equalizers, to emphasize the fact that the formula for θ_0 is essentially similar to the standard equalizer design formula

$$e^{\vartheta} = 1 + \frac{Z_{11}}{R}. \tag{9}$$

The choice of an appropriate Z_{11} therefore requires only routine design methods.[3] The same correspondence with the conventional formula will be found to hold also for most of the design equations to be given later.

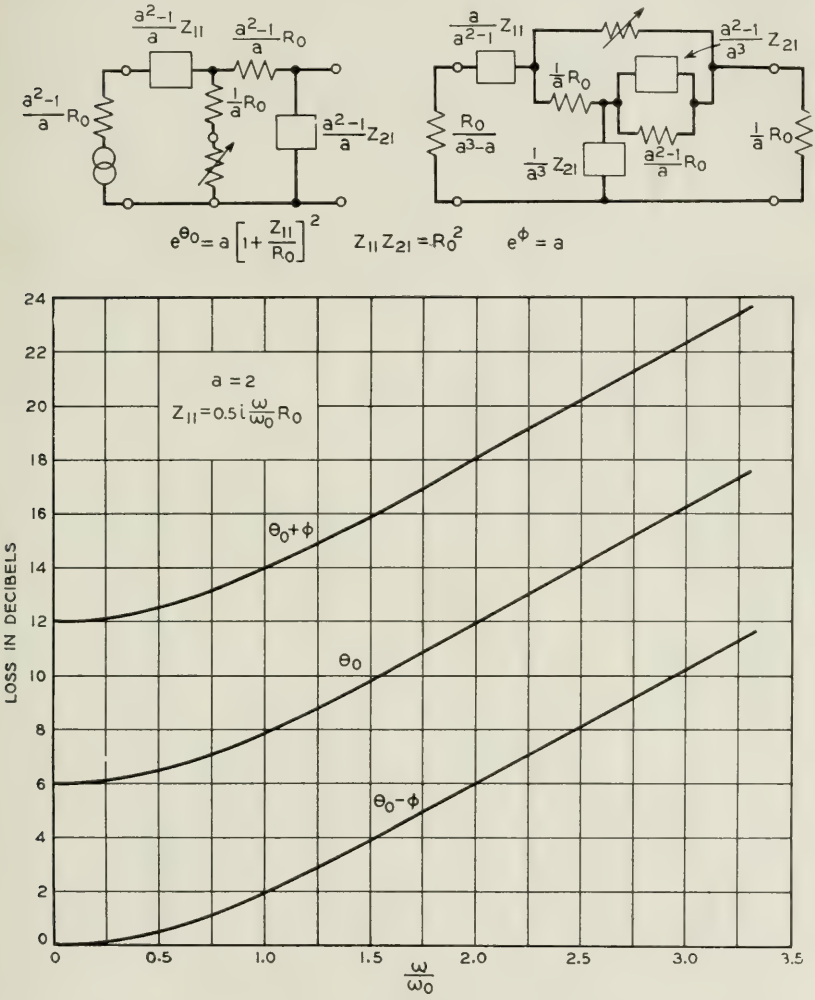


Fig. 8—Variable equalizers with varying reference characteristics.

A third simple network is shown on Fig. 9. Its properties are the converse of those secured from the structures of Fig. 8. The reference loss, θ_0, is now a constant, while φ varies with frequency in a manner

[3] See, for example, "Distortion Correction in Electrical Circuits," O. J. Zobel, *Bell Sys. Tech. Jour.*, July, 1928.

which depends upon the choice of Z. An additional control of ϕ can of course be obtained by the introduction of an auxiliary structure in front of the variable resistance. As in the previous example, the illustrative curves are drawn on the assumption that the characterizing

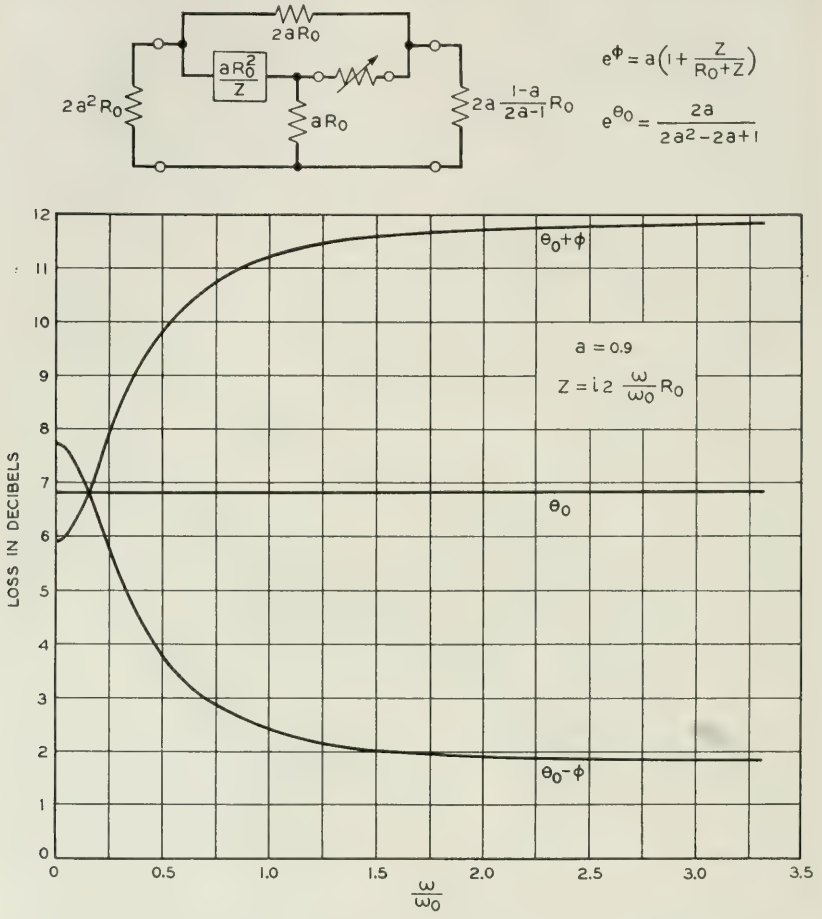
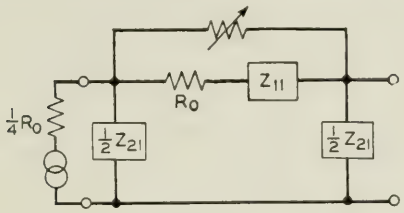


Fig. 9—A variable equalizer requiring only one general impedance branch.

impedance Z is a simple inductance. It will be observed that the curves "see-saw," the attenuation at certain frequencies increasing while that at other frequencies is decreased. This phenomenon depends upon the choice of the parameter a. It disappears when a is assigned either extreme value $\frac{1}{2}$ or 1, and becomes most pronounced at the intermediate value $a = 1/\sqrt{2}$. A similar effect can also be

produced in the networks we have already considered since, as equation (8) shows, the variable attenuation will change sign if the phase shift of the auxiliary network is allowed to increase beyond 45°.

In the fourth structure, shown by Fig. 10, both θ_0 and φ are variable.



$$e^{\Phi} = \left[1 + \frac{Z_{11}}{R_0}\right]$$

$$e^{\Theta_0} = \left[1 + \frac{Z_{11}}{R_0}\right]^2$$

$$Z_{11}Z_{21} = R_0^2$$

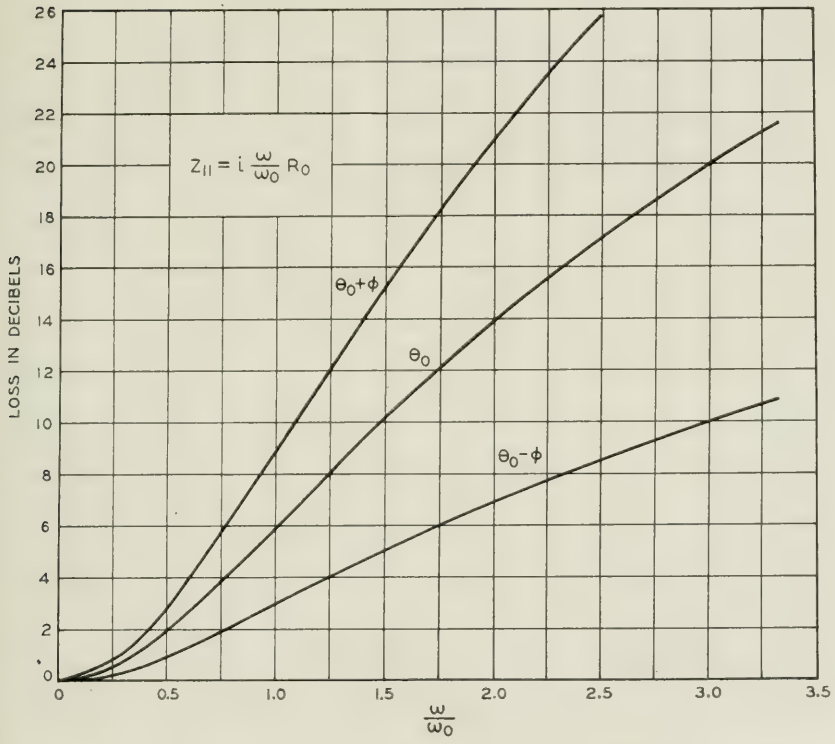
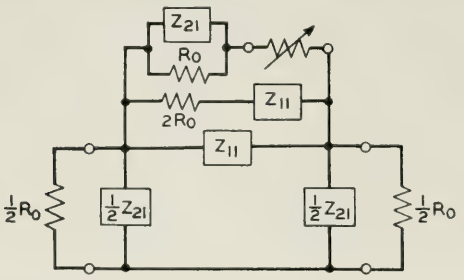


Fig. 10—A variable equalizer with a constant ratio between the reference loss and the variable characteristic.

Their ratio, however, is constant, and the overall loss characteristic of the structure is therefore proportional to some fixed characteristic at all settings of the variable resistance. This property suggests that the network may be of interest for such applications as the equalization of varying lengths of a given transmission line.

The sixth network, shown in Fig. 11, is structurally more elaborate than any which have preceded it. Its design equations are also relatively



$$e^{\phi} = 1 + \frac{1}{2}\frac{\left(\frac{Z_{11}}{R_0}\right)^2}{1+\frac{Z_{11}}{R_0}}$$

$$e^{\theta_0} = \left[1 + \frac{Z_{11}}{R_0}\right]^2$$

$$Z_{11} Z_{21} = R_0^2$$

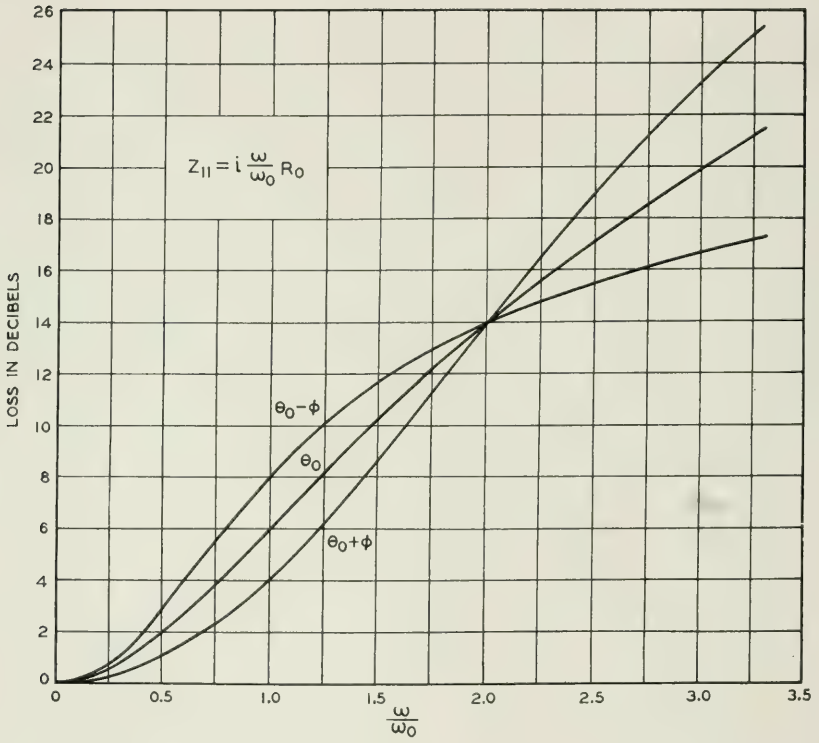
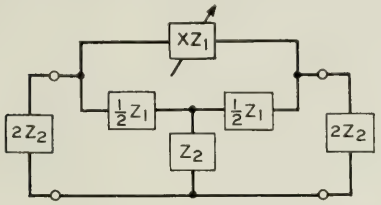


Fig. 11—A variable equalizer with zero loss at one frequency for all settings of the controlling element.

elaborate and difficult to deal with. It possesses, however, the salient property that when $Z_{11} = 0$, perfect transmission of power from one terminating impedance to the other is secured at any setting of the variable resistance. This property suggests that the network may be

useful for systems where it is necessary to introduce variable equalization without attenuation of the channels having the lowest signal level.

The final structure is shown in Fig. 12. Its chief point of interest



$$e^{\phi} = \left[1 + \frac{Z_1}{4Z_2}\right]$$

$$e^{\theta_0} = 2\left[1 + \frac{Z_1}{8Z_2}\right]$$

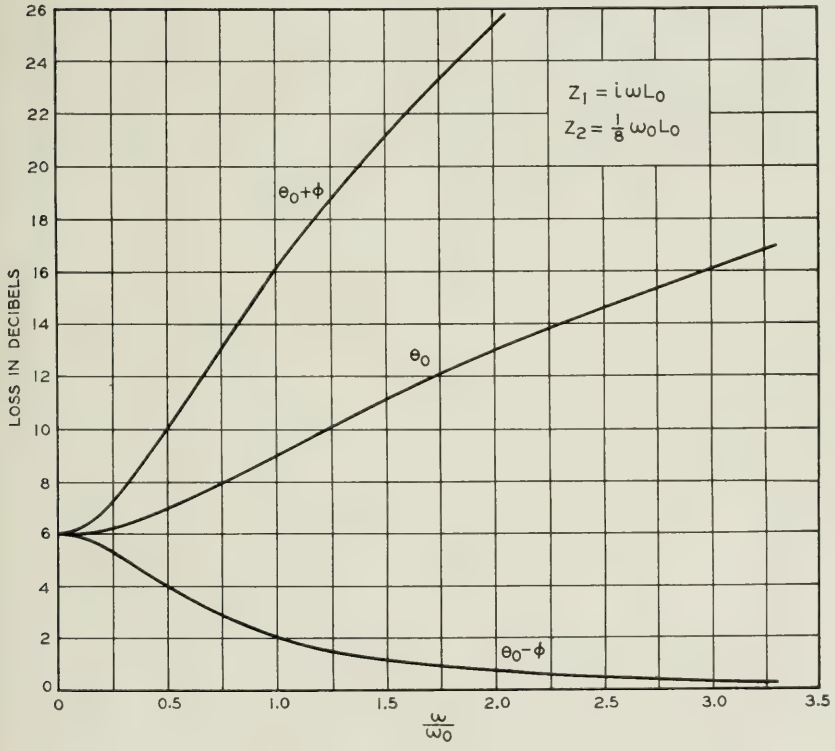


Fig. 12—A variable equalizer adapted to a general controlling impedance.

is the fact that the controlling element, instead of being a variable resistance, is a variable general impedance, which has been labelled xZ_1 in the drawing. In practical cases, of course, Z_1 will ordinarily be a simple resistance, inductance, or capacity. In addition to the variable branch the network also includes two fixed branches propor-

tional to Z_1 and three fixed branches proportional to a second general impedance Z_2. The change from a variable resistance to a variable impedance makes little difference in the analysis. It is merely necessary to replace each R and R_0 by a Z or Z_0 in every equation. So long as equation (3), with the appropriate modification, is satisfied, as it is in this structure, the resulting family of attenuation characteristics will have the same general symmetrical form as those obtained from the resistance controlled devices. A set of curves illustrating this point is shown at the bottom of Fig. 12. They are drawn on the assumption that the Z_1 and Z_2 impedances are respectively inductances and resistances.

The structure of Fig. 12 will still function satisfactorily as a variable equalizer if it is turned inside out so that the $Z_1/2$ impedances become the terminations and the central shunt branch becomes the variable. In this event the present variable impedance must be set at its nominal value. The resulting structure is essentially the inverse of the network of Fig. 12. In the same way, of course, each of the other configurations which have been described can be replaced by its inverse.

An Explanation of the Common Battery Anti-sidetone Subscriber Set

By C. O. GIBBON

THE telephone transmitter serves to convert sound waves into their electrical facsimile; but in performing this primary function the transmitter also acts as an amplifier. Under some conditions the electrical power output of a transmitter may be more than a thousand times as great as the acoustic power activating it. Part of this greatly augmented power is dissipated in the circuit of the telephone set; part is impressed upon the telephone line, whence it is propagated on to the distant listener; and part finds its way into the receiver of the same set, where it is reconverted into sound waves. Speech or noise, picked up by the transmitter and reproduced by the receiver of the same set, is called sidetone.

Noise picked up and amplified by the transmitter and heard as sidetone tends to obscure incoming speech, thereby impairing reception. Similarly, the sound of his own voice, heard more loudly than normal as sidetone because of transmitter amplification, impels the talker involuntarily to lower his voice; thus impairing the reception of his speech at the far end of the connection. The consequent desirability of reducing sidetone has long been recognized, and operator and subscriber sets which accomplish this have been developed.

Circuit schematics of the common battery sidetone and anti-sidetone subscriber sets at present standard in the Bell System are shown in Fig. 1. The anti-sidetone set has become increasingly common during the past few years, and because of the improvements in effective transmission which it affords, bids fair ultimately to be well nigh universally employed. It is, therefore, not surprising that numerous requests have arisen for an explanation of this anti-sidetone circuit which may be more easily followed than one based on the methods usually employed in network analysis. The present paper provides an explanation by means of diagrams with a minimum of mathematical treatment which it is believed those to whom the mathematical approach does not appeal will find helpful in picturing the behavior of this circuit.

The explanation given in this paper is, however, confined to idealized conditions. No concern is given to whether the conditions necessary to exact attainment of the balances described are actually feasible;

nor is any attempt made to discuss questions of practical design beyond pointing out something of the nature of the problems involved. Equations for this anti-sidetone circuit are given and discussed in an appendix, and a vector diagram is shown which illustrates graphically relations among the currents and voltages under the ideal condition of exact balances.

Rearrangement of Circuit Patterns

Simplified explanations of anti-sidetone sets are most frequently based upon analogues with balanced arrangements resembling the

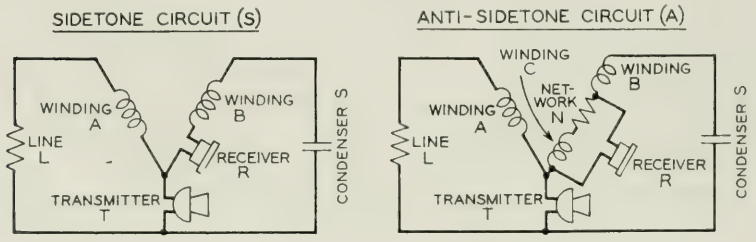


Fig. 1

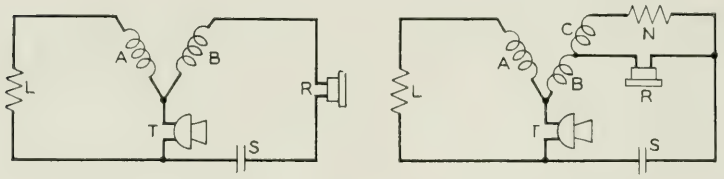


Fig. 2

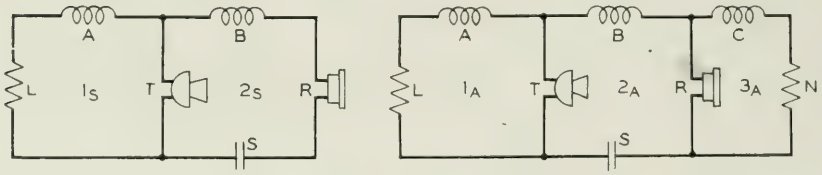


Fig. 3

Group I—Convenient schematic rearrangements

familiar circuit pattern of the Wheatstone bridge, and several such explanations have been devised for the set now to be considered. A wholly different approach will, however, be employed here. The schematics supplementing the present discussion have been arranged in groups to assist in visualizing and coordinating the steps in the explanation. The diagrams of Group I rearrange the familiar conventional schematics of the common battery sidetone and anti-sidetone circuits in Fig. 1 into the patterns most convenient for present purposes

From the intermediate step in Fig. 2 the circuits in Fig. 3 will be seen to be the same electrically as those in Fig. 1. Inasmuch as its effects upon the features to be discussed are negligible, the ringer branch has been omitted. The circuit meshes in Fig. 3 are numbered for identification, subscripts S and A suggestively differentiating meshes of the sidetone and anti-sidetone circuits. Mesh currents and e.m.f.'s in all subsequent schematics are correspondingly identified as later described.

Complementary Sidetone Circuit

From Fig. 3 the three-mesh anti-sidetone circuit is seen to differ schematically from the two-mesh sidetone circuit only by having a balancing branch consisting of a third induction coil winding C and a network N bridged across the receiver to form the third mesh. In this theoretical discussion, N may be assumed to be a network of whatever form is capable of providing the impedance characteristics needed to meet the requirements later discussed; in practice, it becomes merely a resistance integrally comprised within the resistance component of the self-impedance of winding C. The sidetone circuit which would remain if this added branch were disconnected will be called the complementary sidetone circuit or the sidetone complement of the anti-sidetone circuit. In every reference to a sidetone circuit hereafter, the sidetone complement of the associated anti-sidetone circuit is meant.

Receiving and Transmitting Features

The theoretical features of the anti-sidetone circuit to be explained are:

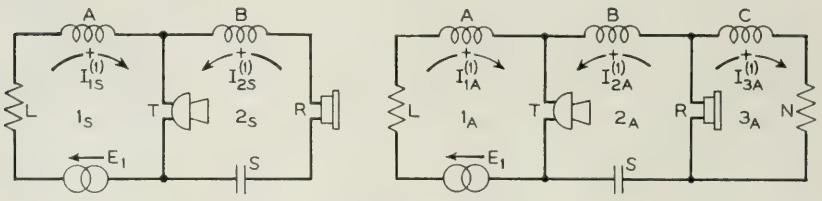
1. The receiving efficiency is the same as that of the complementary sidetone circuit.
2. The transmitting efficiency is the same as that of the complementary sidetone circuit.
3. The current through the receiver when transmitting, i.e., the sidetone, is zero.

The above efficiencies, it should be noted, are purely circuit efficiencies; they do not include electro-acoustic conversions by the instruments, and they should not be confused with questions pertaining to effective transmission performance.

Circuit Conditions When Receiving and When Transmitting

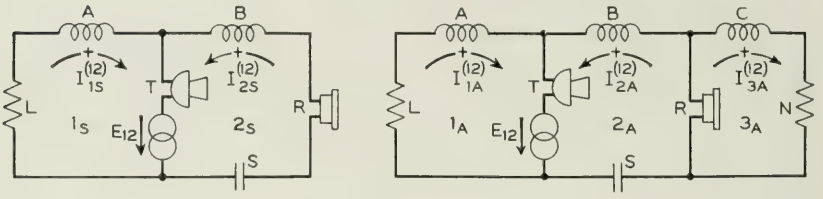
The following discussion of the diagrams in Groups II and III applies to both the sidetone and the anti-sidetone circuits.

Group II shows the e.m.f.'s impressed upon the circuits when receiving and when transmitting, and the assumed positive directions of the resulting mesh currents.[1] The schematics in Fig. 4 will be used in the later explanations of the receiving condition; but for the transmitting condition the equivalent component schematics evolved in Group III will be employed instead of Fig. 5.



Receiving

Fig. 4



Transmitting

Fig. 5

Group II—The applied E.M.F.'s and assumed * mesh currents, receiving and transmitting

By the well-known principle called Thevenin's Theorem, under the receiving condition the line and the set at its distant end may be replaced by an impedanceless generator in series with the line impedance, the e.m.f. of this generator being equal to the open-circuit voltage across the terminals of the line. Also, under receiving conditions, the transmitter will be treated as a passive impedance; and

* The curved arrows in the above and in all subsequent circuit schematics, it should be noted, do not purport to indicate the relative directions of the actual currents; but merely to denote the directions *assumed* as the positive sense of mesh currents and voltages. The selection of these directions is purely optional. As a matter of convenience, they have been so chosen that the currents through the transmitter and receiver are in every case the algebraic sum of the two contributing mesh currents.

[1] The following conventions and nomenclature are employed, see Group II:— The directions around the meshes, chosen as positive for all voltages and currents, are indicated by curved arrows. Subscripts differentiate impressed e.m.f.'s with regard to the meshes in which they are applied, and identify currents with respect to the meshes (or circuit elements—see Group IV) in which they flow. Currents are further identified with respect to the e.m.f. or e.m.f.'s to which they are due, by superscripts corresponding with the activating voltage subscripts.

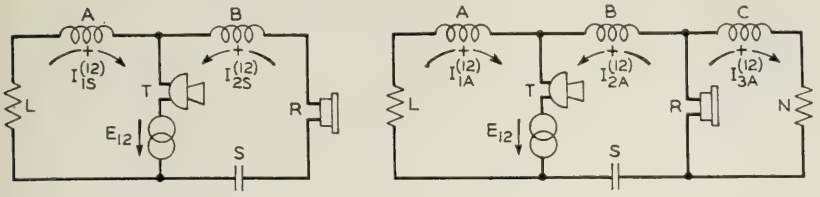


Fig. 5

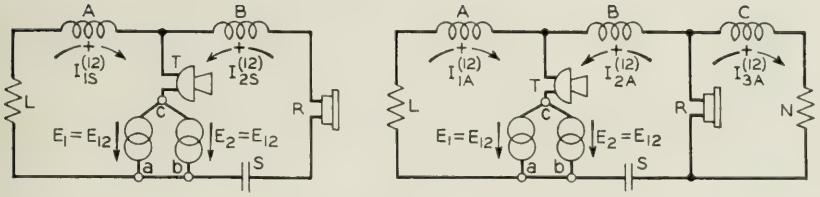


Fig. 6

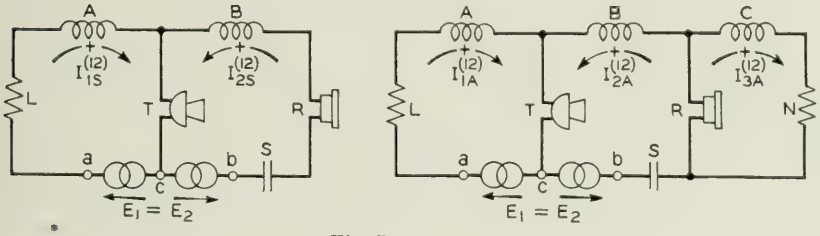


Fig. 7

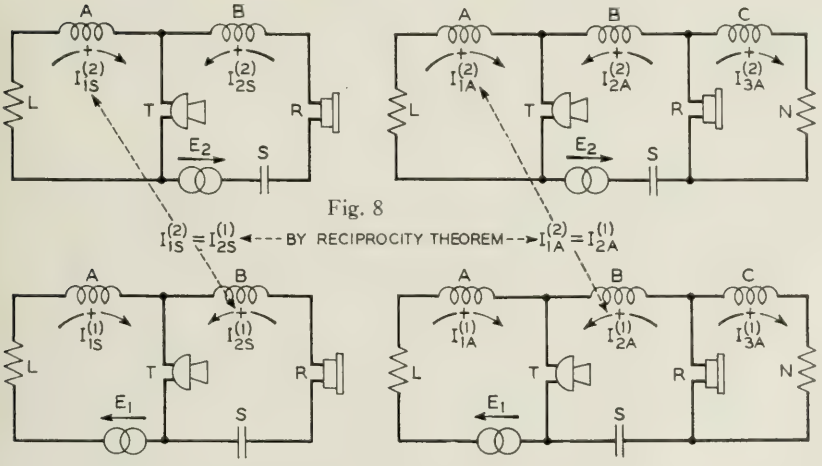


Fig. 8

Fig. 9

Group III—Components of transmitting E.M.F.'s and assumed mesh currents

when transmitting (see Fig. 5 of Group III), it will be looked upon as the same passive impedance in tandem with an impedanceless generator whose e.m.f. equals the variations in voltage drop (of the battery supply current) across the transmitter due to the changes in its impedance which occur when the transmitter is agitated by sound. This e.m.f. thus replaces in the circuit the sound engendered variations in the transmitter impedance, thereby permitting this impedance to be treated as a constant. Being impedanceless, this generator may, without effect upon the circuit, be replaced by two impedanceless generators—each having the same e.m.f. as the first—connected in parallel as shown in Fig. 6. The direct connection between points *a* and *b*, however, is shunted by the impedanceless path *acb*, so that the direct connection *ab* may, without effect, be broken as in Fig. 7. Hence, the two equal e.m.f.'s in Fig. 7, acting simultaneously, are equivalent to the single e.m.f. in Fig. 5; and the mesh currents in the two figures are, therefore, identical. Hereafter, Fig. 7, rather than Fig. 5, will be considered the transmitting condition.

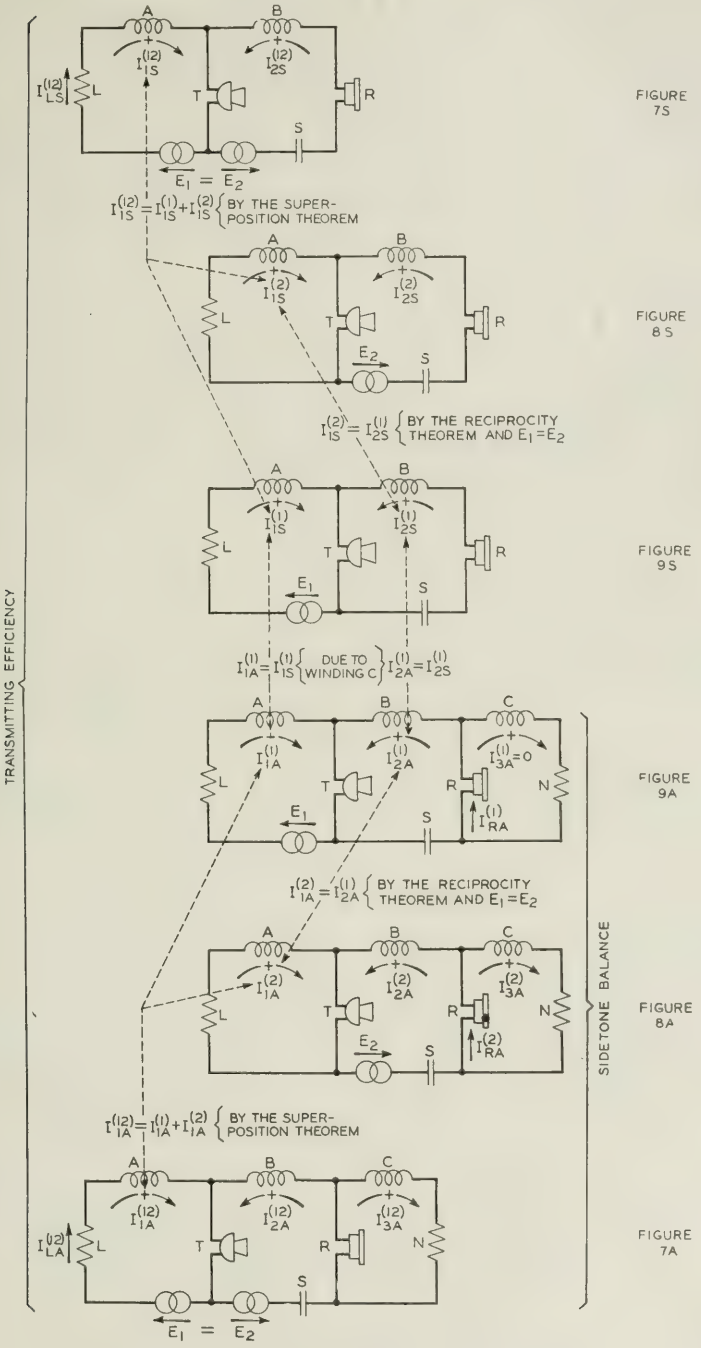
This transmitting condition may, however, be broken into two component conditions. By the fundamental principle known as the Superposition Theorem, the currents in Fig. 7 are equal to the sum of the currents which would result from each of the two e.m.f.'s acting alone. In other words, the transmitting currents in Fig. 7 are equal to the sum of the corresponding currents in Figs. 8 and 9. But by a second fundamental principle called the Reciprocity Theorem, the current at any point X in a circuit, due to an e.m.f. at any other point Y, is equal to the current which would result at Y from an equal e.m.f. at X. Applying this to Figs. 8 and 9, in which $E_1 = E_2$, the mesh currents pointed out by the arrows joining these two schematics are equal, viz.:

$$I_{1S}^{(2)} = I_{2S}^{(1)} \qquad \text{and} \qquad I_{1A}^{(2)} = I_{2A}^{(1)}. \tag{1}$$

Of the above components of the transmitting currents, those in Fig. 9 are due to an e.m.f. acting in mesh 1, i.e., in series with the line impedance. This, however, is also the condition when receiving, as will be seen by comparing Figs. 9 and 4.

Neutralizing Balance—Receiving Efficiency

Consider next the purpose of winding *C*. It is, of course, desirable that the transmitting and receiving efficiencies be undiminished by the anti-sidetone arrangement. If it is possible so to adjust the couplings among windings *A*, *B* and *C* that the current $I_{3A}^{(1)}$ in Fig. 4 or 9 is zero, the balancing branch can then be disconnected without effect



Group IV—Transmitting efficiency—sidetone balance

upon the currents, and the circuit will thereby be reduced to the sidetone circuit. Hence, with this ideal adjustment of the couplings, the receiving efficiency of the anti-sidetone circuit will be the same as that of its sidetone complement.

Although equally satisfactory designs could be worked out with other polings of the coil windings, the relative inductive directions among windings A, B and C in the circuit here dealt with are such that, if a current were passed through all three in series, windings A and B would be inductively aiding; and C would be inductively opposed to both A and B. Returning to the above condition for maintaining the receiving efficiency, namely, that $I_{3A}^{(1)}$ be made zero, the windings of the coil must be so adjusted that the sum of the two voltages induced in winding C by its inductive couplings with windings A and B is equal and opposite the voltage drop across the receiver. With the windings poled as just stated, this requires that

$$(+ I_{1A}^{(1)} Z_{AC}) + (- I_{2A}^{(1)} Z_{BC}) = - (- I_{2A}^{(1)} Z_R). \tag{2}$$

This voltage balance expressed by eq. (2) will be referred to as *neutralizing balance:* its attainment requires the coil windings be adjusted to meet the relation shown by eq. (6) in the appendix.

It is important to note that neutralizing balance, and hence the efficiency relations which depend upon it, are independent of the impressed e.m.f. E_1, of the line impedance Z_L and the self-impedance of winding A, and of the network impedance Z_N and the self-impedance of winding C. This of course follows from the fact that none of these quantities is involved in eq. (2).

Transmitting Efficiency

It will now be shown that the transmitting efficiency of the anti-sidetone circuit is the same as that of its sidetone complement; and that this equality, like that of the receiving efficiencies, results from the neutralizing balance effected by winding C. This is true if, with equal transmitter e.m.f.'s in the top and bottom diagrams of Group IV, the line currents are equal, viz., if

$$I_{1A}^{(12)} = I_{1S}^{(12)}.$$

To prove this relation, refer to Fig. 7A at the bottom of Group IV and move up step by step to Fig. 9A, observing the relations between mesh currents indicated by the arrows. It will be seen that

$$I_{1A}^{(12)} = I_{1A}^{(1)} + I_{2A}^{(1)}.$$

But in Figs. 9*A* and 9*S*, due to neutralizing balance, as was shown in discussing receiving efficiencies,

$$I_{1A}^{(1)} = I_{1S}^{(1)} \qquad \text{and} \qquad I_{2A}^{(1)} = I_{2S}^{(1)}.$$

Finally, continuing from Fig. 9*S* upward to Fig. 7*S*, it is seen that

$$I_{1S}^{(1)} + I_{2S}^{(1)} = I_{1S}^{(12)}.$$

Hence,

$$I_{1A}^{(12)} = I_{1A}^{(1)} + I_{2A}^{(1)} = I_{1S}^{(1)} + I_{2S}^{(1)} = I_{1S}^{(12)}.$$

Primary Purposes of Winding *C* and of Neutralizing Balance

The above relations between the efficiencies of the anti-sidetone circuit and those of its sidetone complement are, however, merely incidental to the primary purposes of winding *C* and of the neutralizing balance which it provides. The major purpose of winding *C* is that, entirely apart from its neutralizing action, the voltages induced in it through its couplings make it possible to obtain sidetone balance by adjusting Z_N; i.e.—referring to Fig. 7*A* at the bottom of Group IV—given any value of Z_L, it is theoretically possible so to adjust Z_N that $I_{2A}^{(12)} = -I_{3A}^{(12)}$. The current through the receiver under the transmitting condition, i.e., sidetone, will then be zero. Neutralizing balance permits this adjustment of Z_N to be made without affecting the circuit efficiencies.

Sidetone Balance

The following discussion of sidetone balance will proceed on the assumption that the couplings of winding *C* with windings *A* and *B* have already been adjusted for neutralizing balance, since this condition is required to maintain the circuit efficiencies. This approach is merely a matter of convenience, however, for it will be indicated that the impedance of *N* required to effect sidetone balance is the same whether the neutralizing balance is taken into account or ignored. Although sidetone balance is made possible by the couplings of winding *C*, and the impedance of *N* needed to reduce sidetone to zero does depend upon the values to which the self and mutual impedances of this winding have been adjusted, neither the attainment of sidetone balance nor the value of Z_N required to provide it depends upon the existence of neutralizing balance.

If sidetone is to be zero, the voltage across the receiver under the transmitting condition, i.e., the sum of the voltages across *C* and *N*, must be made zero. Expressed in terms of the voltages in Fig. 7*A* at the bottom of Group IV, this requires that

$$I_{1A}^{(12)} Z_{AC} - I_{2A}^{(12)} Z_{BC} - I_{3A}^{(12)} (Z_C + Z_N) = 0. \qquad (3)$$

Here, at once, the dependence of sidetone balance upon the presence of the inductively coupled third winding is apparent. Without winding C the impedances Z_{AC}, Z_{BC} and Z_C in the above expression would all be zero, the terms in which they occur would drop out, and the requirement for elimination of sidetone would reduce to $Z_N = 0$, i.e., a short circuit across the receiver.

The remainder of this discussion of sidetone balance can be carried out more conveniently in terms of the mesh currents than in terms of the above voltages. As already noted, the voltage balance just examined is equivalent to requiring that $I_{2A}^{(12)}$ and $I_{3A}^{(12)}$ in Fig. 7A be made equal and opposite. But $I_{2A}^{(12)}$ is the sum of the two components, $I_{2A}^{(1)}$ in Fig. 9A and $I_{2A}^{(2)}$ in Fig. 8A; and, because of neutralizing balance, $I_{3A}^{(12)} = I_{3A}^{(2)}$ in Fig. 8A. Furthermore, by the Reciprocity Theorem, the component $I_{2A}^{(1)}$ always equals $I_{1A}^{(2)}$; and the latter, like the former, is independent of Z_N. The condition for sidetone balance may, therefore, be expressed in terms of the currents in Fig. 8A as

$$I_{1A}^{(2)} + I_{2A}^{(2)} + I_{3A}^{(2)} = 0. \tag{4}$$

The question, then, is whether N can be so adjusted that the sum of the three mesh currents in Fig. 8A is made zero; and it is fairly evident such an adjustment for any given value of Z_L is theoretically possible. Since $I_{1A}^{(2)}$ is known to be independent of Z_N, and because inspection shows the circuit to be symmetrical with respect to L and N, it appears that $I_{3A}^{(2)}$ must be independent of L—an intuitive inference which the Reciprocity Theorem confirms. The value of $I_{2A}^{(2)}$ depends, of course, upon both Z_L and Z_N. Hence, with the value of $I_{1A}^{(2)}$ remaining fixed as Z_N is varied, and with $I_{3A}^{(2)}$ independent of Z_L but under the direct control of Z_N, it may be concluded possible to meet eq. (4) by a suitable choice of Z_N for any given value of Z_L. The value of Z_N required to attain sidetone balance is shown by eq. (7) in the appendix.

With N so adjusted that sidetone is zero, it is obvious the receiver impedance may be changed in any way whatever without upsetting the sidetone balance. The same is true of any change in the impedance of the transmitter; because this, being equivalent to a compensating change in the transmitter e.m.f., would cause all mesh currents to change in the same proportion; thus leaving the balance expressed by eq. (4) undisturbed. Hence, the impedance of N required to provide sidetone balance is independent of the receiver and of the transmitter. But as has already been seen, the couplings of winding C necessary to provide the neutralizing balance in eq. (2) do depend upon the receiver and transmitter impedances. The significance of this observation is that although the value of Z_N required

to provide sidetone balance does depend upon the values of the couplings of C, neither the attainment of the balance in eq. (4) nor the impedance of N required to provide it depends upon the balance in eq. (2) being met. In other words, the neutralizing balance expressed by eq. (2), and the sidetone balance expressed by eq. (4), are mutually independent; either may be attained without the other.

Practical Considerations

With the simple types of coil and network permitted by economic and space limitations, the balances upon which the above performance of the anti-sidetone circuit depends can be obtained exactly only with a given line and at a single frequency. For practical purposes, however, exact balances are needless. Sound leakage under the receiver cap and conduction through the head structure fix a limit beyond which further reduction in sidetone is not of value. Actual designs, therefore, aim at the best compromise in reducing sidetone over the voice range and the range of line impedances important in practice, as judged by the resulting effective transmission performance obtained with the instruments employed. Under typical plant conditions, designs now in service reduce the volume of sidetone with present instruments to a level averaging around 10 to 12 db below that of the complementary sidetone sets.

APPENDIX

Algebraic Solution of Circuit Equations

Referring to Fig. 10, the following circuit equations may be written:

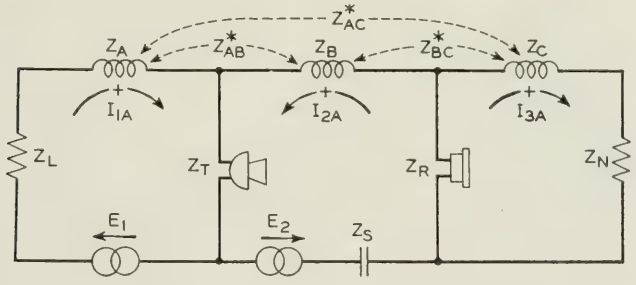
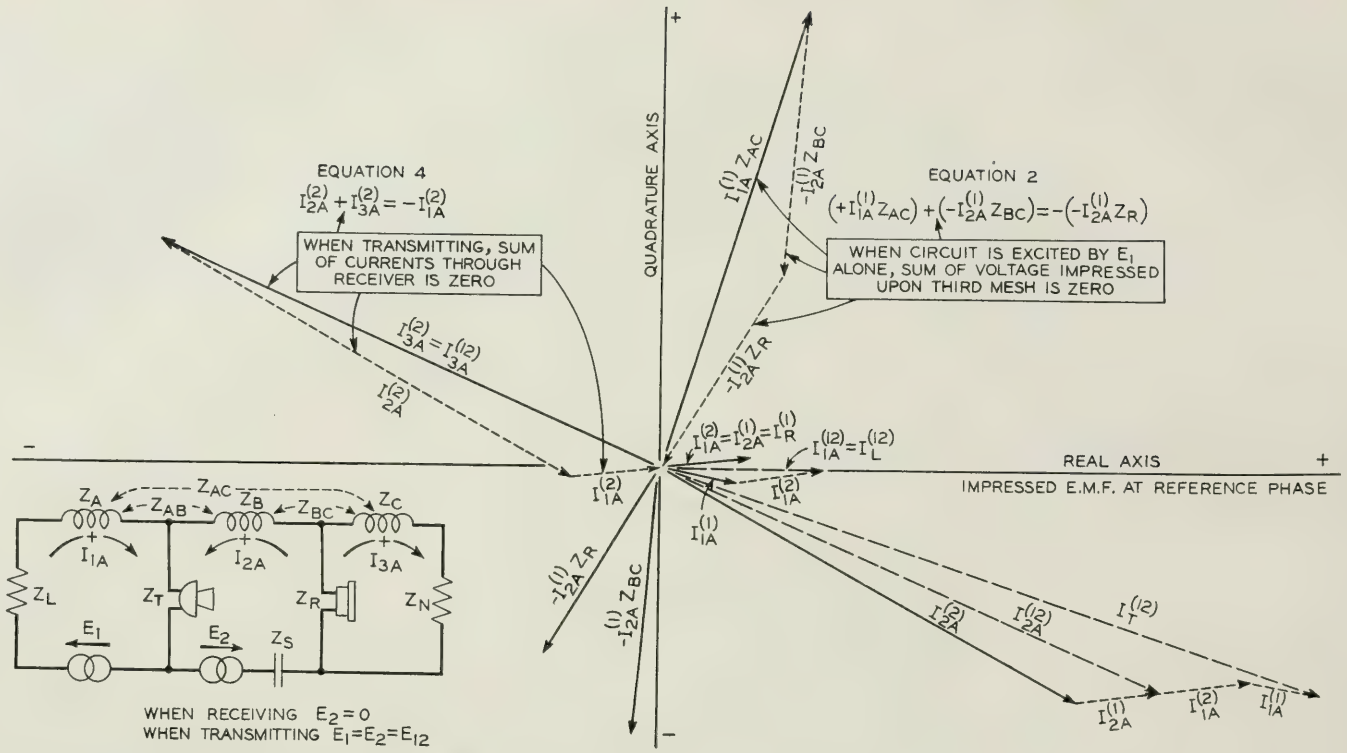


Fig. 10—* The poling of windings A and B is series aiding. Winding C is poled in series opposition to windings A and B.

$$\left.\begin{aligned} (Z_L+Z_A+Z_T)I_{1A}+\quad & (Z_T-Z_{AB})I_{2A}-\quad & Z_{AC}I_{3A}=E_1 \\ (Z_T-Z_{AB})I_{1A}+(Z_T+Z_B+Z_R+Z_S)I_{2A}+\quad & (Z_R+Z_{BC})I_{3A}=E_2 \\ -Z_{AC}I_{1A}+\quad & (Z_R+Z_{BC})I_{2A}+(Z_R+Z_C+Z_N)I_{3A}=0 \end{aligned}\right\}. \tag{5}$$


EQUATION 4
$I_{2A}^{(2)} + I_{3A}^{(2)} = -I_{1A}^{(2)}$
WHEN TRANSMITTING, SUM OF CURRENTS THROUGH RECEIVER IS ZERO
EQUATION 2
$(+I_{1A}^{(1)}Z_{AC}) + (-I_{2A}^{(1)}Z_{BC}) = -(-I_{2A}^{(1)}Z_R)$
WHEN CIRCUIT IS EXCITED BY E_1 ALONE, SUM OF VOLTAGE IMPRESSED UPON THIRD MESH IS ZERO
QUADRATURE AXIS
REAL AXIS
IMPRESSED E.M.F. AT REFERENCE PHASE
Z_A Z_{AC} Z_{AB} Z_B Z_{BC} Z_C
I_{1A} I_{2A} I_{3A}
Z_L Z_T Z_R Z_N
E_1 E_2 Z_S
WHEN RECEIVING $E_2 = 0$
WHEN TRANSMITTING $E_1 = E_2 = E_{12}$

This diagram shows the following:

1. Vectors of mesh currents due to E_1 acting alone.
2. Vectors of mesh currents due to E_2 acting alone.
3. Vector additions of the foregoing components showing the following under the transmitting condition:
   - A. Vectors of the mesh currents.
   - B. Vectors of the currents through the transmitter and fed into the line.
   - C. The sum of the currents through the receiver is zero.
4. Vectors of voltages impressed upon the third mesh when circuit is excited by E_1 alone.
5. Vector addition of foregoing voltages showing sum to be zero.

Fig. 11—Vector diagram for assumed example of anti-sidetone common battery subscriber set circuit

These cover both transmitting and receiving conditions: when transmitting, $E_1 = E_2$; and when receiving, $E_2 = 0$.

The relation which the induction coil must meet in order to provide neutralizing balance can be determined by solving eqs. (5) under the receiving condition $E_2 = 0$, and imposing the requirement that $I_{3A}^{(1)} = 0$. This gives as the relation to be met.

$$\frac{Z_{AC}}{Z_{BC} + Z_R} = \frac{Z_{AB} - Z_T}{Z_T + Z_B + Z_R + Z_S}. \tag{6}$$

In like manner, the value of Z_N needed to provide sidetone balance can be determined by solving eqs. (5) under the transmitting condition $E_1 = E_2$, and imposing the requirement that $I_{2A}^{(12)} + I_{3A}^{(12)} = 0$. This value of Z_N, regardless of whether or not eq. (2) is imposed as a further condition in its derivation, is found to be

$$Z_N = Z_{BC} - Z_C + \frac{Z_{AC}(Z_{AC} + Z_{BC} - Z_{AB} - Z_B - Z_S)}{Z_L + Z_A + Z_{AB}}. \tag{7}$$

Note that Z_N is here independent of Z_T and Z_R, except as these may enter implicitly as factors affecting the impedances at right in designing the coil for optimum performance with specified instruments. In other words, the transmitter and receiver may be changed without disturbing the sidetone balance. Such a change would, however, upset the neutralizing balance, thereby altering the efficiencies from those of the sidetone circuit.

Vector Diagram

Relations among the component mesh currents in an anti-sidetone circuit of this type under ideal conditions of exact neutralizing and sidetone balances, are illustrated by the vector diagram in Fig. 11. As all of the current vectors indicate current per volt impressed, those for the mesh currents under the receiving condition in Fig. 4*A* are identical with those under the component of the transmitting condition in Fig. 9*A*. Vector sums of the mesh currents show the current through the receiver and that fed into the line when transmitting, and illustrate the sidetone balance. Vectors of the three voltages acting around the third mesh in Figs. 4*A* and 9*A* are also shown, together with their summation. The latter illustrates the neutralizing balance of eq. (2).

The Occurrence and Effect of Lockout Occasioned by Two Echo Suppressors

By ARTHUR W. HORTON, Jr.

"The Time Factor in Telephone Transmission" by O. B. Blackwell (B. S. T. J. January 1932) deals with a number of problems which arise in connection with telephone circuits having long transmission times. This paper discusses one such effect, the occurrence of lockout caused by the echo suppressors involved in a long telephone connection.

The occurrence of lockout is shown to cause an increase in repetition rate, which is ordinarily small for circuits as now used commercially. The increase in repetition rate is approximately proportional to the number of lockouts occurring and to their mean duration, or to the per cent of time locked out.

The expected number of lockouts is shown to depend upon the characteristic time intervals of conversational speech, the relay hangovers, the delay of the circuit and location of the echo suppressors with respect to the ends of the circuit. Subject to certain restrictions, the expected number of lockouts increases with the delay included between the echo suppressors, and is nearly independent of the delays between the suppressors and the circuit terminals.

The mean duration of lockouts is shown to be proportional to the relay hangovers.

Introduction

WHEN carrying on a conversation over a telephone circuit of moderate length, the subscribers are ordinarily unaware of any limitations imposed upon the free interchange of information. As the length of the circuit is increased the time factor [1] becomes increasingly important and may become manifest in a number of ways. One result of the time factor is the occurrence of echoes which become apparent when the speech energy reflected from the end of the circuit is delayed in returning to the talking subscriber. When the circuit is equipped with an echo suppressor to render this effect unnoticeable, or when a long connection of two such circuits is made, the action of the suppressors is such as to make the circuit inoperative in the opposite direction to which speech is being transmitted. Consequently the subscribers are no longer able to interchange information with the ease and rapidity that would be enjoyed on a shorter circuit.

[1] "The Time Factor in Telephone Transmission," O. B. Blackwell, *Bell System Technical Journal*, January 1932.

A circuit equipped with a single echo suppressor is always operative in one direction, and although both subscribers may start to talk at about the same instant, one or the other will always obtain control of the circuit and his speech will be heard by the other subscriber. The principal difficulties encountered on circuits of this type become apparent when the hangover times of the relays are large. There is some difficulty in interrupting since the relays do not release during the pauses between words, and a quick response following a pause by the first talker may reach the suppressing relay before it has released, resulting in a mutilation of the initial part of the response.

When two echo suppressors are used, as is the case when two circuits each equipped with an echo suppressor are connected in tandem, similar difficulties may be encountered. In addition, lockout, or blocking of transmission in both directions, may occur and may persist for an appreciable time. Since neither subscriber is aware that the other is talking, both may continue talking until one or the other of the relays releases during a pause and enables the circuit in the appropriate direction. Thus neither subscriber will be conscious of the fact that a lockout has occurred unless he realizes from the context that some part of the conversation has been lost.

This paper discusses the manner in which lockouts can occur, and presents the results of a series of tests to determine their effect upon conversation as measured by repetition rate.[2] These results indicate that the repetition rate increases with the per cent of time during which lockout occurs. It is shown that the locked out time can be approximately calculated in terms of the circuit constants and suitable characteristic intervals of conversational speech, and the calculated values can in turn be used to predict the effects of lockout on repetition rate.

In terms of the effect upon the talkers, a lockout may be considered to occur when speech currents from one talker are prevented from reaching the other talker by one of the suppressing relays and those same speech currents operate another suppressor in such a way that speech currents from the latter talker are prevented from reaching the former. This description of lockout should not be considered as a precise definition since it does not specify the duration of a lockout. No definition in terms of measurements made upon speech at the circuit terminals would be free from difficulties in practical application, such as that of determining with sufficient precision the instants at which speech is considered to start and stop, and that of determining the direction of transmission. A definition in terms of the operations of

[2] "Rating the Transmission Performance of Telephone Circuits," W. H. Martin, *Bell System Technical Journal*, January 1931.

the suppressors is somewhat simpler to formulate, but may be difficult to apply when the echo suppressors are separated geographically. In the tests to be described there was no such separation involved and consequently the operation of the suppressors could be readily observed and easily and accurately measured. Accordingly for the purpose of this paper we shall define a lockout as the condition in which the suppressors are operated in such a way that both directions of transmission are simultaneously blocked. In general, lockouts may be caused by speech, or noise, or both, but the term will be used here to apply to the case in which operations of the suppressors have been caused by speech from both ends of the circuit.

In the course of a conversation the interchange of speech is ordinarily such that the circuit is alternately disabled by the two suppressors in one direction or the other depending upon the direction of transmission. When a pause of sufficient duration occurs, the party not in control of the circuit may reply at such a time that he obtains control of the echo suppressor nearest to his end of the circuit, and a lockout can occur provided that his speech does not reach the distant suppressor until after the party formerly in control of the circuit has resumed talking and has obtained control of that suppressor. The occurrence of lockout is therefore dependent upon the time intervals in conversational speech and upon the constants of the circuit.

The Manner in Which Lockout can Occur

The characteristic time intervals of conversational speech upon which the occurrence of lockout depends, are treated in a companion paper by Mr. Norwine and Mr. Murphy.[3] It is sufficient here to define two such characteristic intervals based on a simplified concept of a conversation. Neglecting grammatical considerations we can consider speech to be composed of a sequence of vocal intervals defined and separated by silent intervals. The lengths of these silent intervals will be called resumption times. Likewise a conversation may be considered to be composed of an alternate succession of speeches, defined and separated by intervals, the lengths of which will be called response times. An ambiguity occurs when both parties talk simultaneously but, for the purpose of this discussion, it will be sufficient to allow for this situation by admitting negative response times.

Figure 1 represents a generalized four-wire circuit equipped with two echo suppressors located at different distances from the ends of the circuit. The transmission times of the different parts of the circuit are

[3] "Characteristic Time Intervals in Telephonic Conversation," A. C. Norwine and O. J. Murphy, this issue of the *Bell System Technical Journal*.

indicated on the figure with appropriate subscripts and the two directions of transmission are differentiated by the primed and unprimed notation. The suppression points are indicated by arrows which represent an opening of the transmission path when the relays, or other sup-

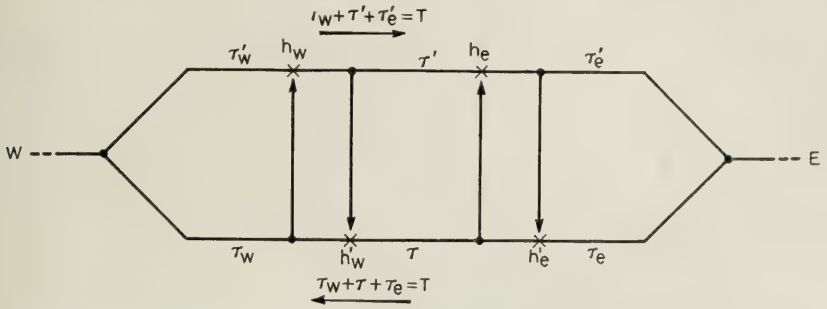


Fig. 1—Schematic of generalized four-wire circuit equipped with two echo suppressors.

pression devices are operated. The suppressing relays are specified by a notation which refers either to the particular relay or to its hangover, or releasing time. According to the definition given above a lockout exists during the time that both the relays h_e and h_w' are operated.[4]

With the exception of the beginning and end of the conversation the occurrence of lockout can be described in terms of the resumption and response times following a pause by one talker, and the constants of the circuit. Referring to Fig. 1, and considering the sequence of events following a pause by E, we shall see that two types of lockout can occur.

The first type, which is the one usually met in practice, can occur when $h_e < h_w + \tau$, and h_w releases after h_e. A response by W and a resumption by E are necessary to produce a lockout. It will persist as long as both E and W continue to talk and for an additional time equal to the delay from the end of the circuit to the first relay to release after a pause by one talker, plus the hangover time of that relay. A lockout of this type may be termed a lasting lockout.

The second type can occur when $h_w + \tau < h_e$, and h_w releases before h_e. It is possible for a response by W to arrive at h_w and operate h_w' before h_e has released thus causing a lockout wihch will be terminated when h_e releases. A lockout of this type, which may be termed a releasing lockout, can occur without a resumption by E, or if E's resumption reaches h_e' after h_e releases. If a releasing lockout has oc-

[4] Also, according to the definition, when both the relays h_w and h_e' are operated, a condition of no practical importance.

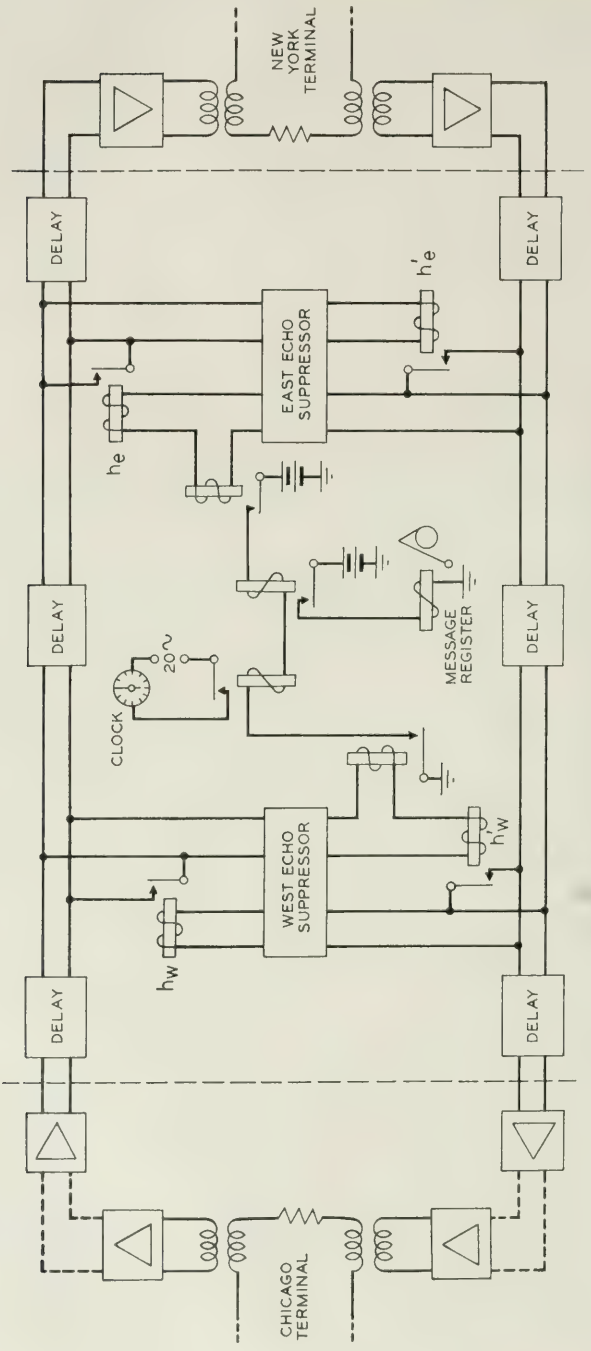


Fig. 2—Circuit arrangement used for experimental results.

curred and E's resumption operates h_e before W's response can operate h_e', a second lockout which will be of the lasting type, will at once occur. Otherwise W's response will operate h_e' giving control of the circuit to W.

Experimental Conditions and Data

To obtain experimental data of the occurrence of lockout in long distance conversations and to determine the resulting effect on repetition rate, added delay and echo suppressors were inserted at the New York end of a circuit to Chicago, Illinois. This circuit is used as a tie line by the Western Electric Company for the transaction of company business between its Hawthorne plant and New York office. The regular echo suppressor usually associated with the circuit at Pittsburgh was removed for these tests. The circuit arrangement employed is shown schematically in Fig. 2, the added equipment being included between the dotted lines. This equipment was adjusted to have zero insertion loss and the frequency characteristic was equalized to within $\pm$ 2 db from 200 to 3000 cycles. The overall net loss from toll board to toll board was 7 db. The suppressors were 44-A echo suppressors operating at a sensitivity of 31 db referred to the zero level point of the circuit, except in those cases specifically mentioned. The added delay circuits were of the acoustic type consisting essentially of a suitable length of brass pipe terminated by high quality loud speaking telephones together with the necessary amplifiers and equalizers to give zero loss over the frequency range from 200 to 3000 cycles. These delay circuits were available in units of 0.023, 0.05, 0.08, 0.10 and 0.15 second, and various combinations of these delays were used together with the tie line delay of 0.043 second to obtain the circuit conditions which were tested.

The details of the recording mechanism are indicated schematically in Fig. 2. A relay was added in series with the shorting relay of each echo suppressor, so that every operation of the echo suppressor relay was accompanied by an operation of the added relay. The simultaneous operation of these relays energized two other relays, one of which in turn operated a message register to record the number of lockouts, and the other connected a 20-cycle oscillator to a cycle counter to record the locked out time.

Service observers at New York monitored both directions of the conversation and recorded repetitions and other pertinent data regarding each call.

The circuit conditions tested are shown in Table I, the notation of which corresponds to Fig. 1. For each condition the first line re-

fers to transmission from west to east and the second line from east to west. The hangovers are those of the relays which short the indicated transmission path, for example, the figure 0.186 in the first line of Table I refers to the hangover of the relay at the west end of the circuit which shorts the transmission path from west to east. The designation in Table I indicates the grouping of conditions for observation. All conditions having the same numeral in the designation were observed concurrently, the procedure being to observe 25 calls on condition *a*, then 25 calls on condition *b* and so on. In this way seasonal variations and uncontrolled effects at the terminals or in the transmission line have been minimized for a group of conditions bearing the same numerical desig-

TABLE I

Condition	T	τ_w	h_w	τ	h_e	τ_e
1*a*	0.139 0.139	0.043 0.043	0.186 0.200	0.073 0.073	0.146 0.146	0.023 0.023
1*b*	0.193 0.193	0.043 0.043	0.186 0.200	0.100 0.100	0.200 0.200	0.050 0.050
1*c*	0.293 0.293	0.043 0.043	0.186 0.200	0.150 0.150	0.300 0.300	0.100 0.100
2*a*	Same as 1*a*					
2*b*	0.139 0.139	0.043 0.043	0.186 0.200	0.073 0.073	0.200 0.200	0.023 0.023
2*c*	0.139 0.139	0.043 0.043	0.186 0.200	0.073 0.073	0.300 0.300	0.023 0.023
3*a*	Same as 1*c*					
3*b*	0.316 0.316	0.043 0.043	0.186	0.250 0.250	0.146	0.023 0.023
4*a*	Same as 1*c*, suppressor sensitivities 28 db					
4*b*	Same as 1*c*, suppressor sensitivities 31 db					
4*c*	Same as 1*c*, suppressor sensitivities 34 db					
5*a*	Same as 1*c*, suppressor sensitivities 31 db					
5*b*	Same as 1*c*, suppressor sensitivities 41 db					
6*a*	Same as 1*c*, suppressor sensitivities 47 db					
7*a*	0.116 0.116	0.043 0.043	0.136 0.150	0.050 0.050	0.146 0.100	0.023 0.023
7*b*	0.116 0.116	0.043 0.043	0.136 0.170	0.050 0.050	0.210 0.146	0.023 0.023

TABLE I (*Continued*)

Condition	T	τ_w	h_w	τ	h_e	τ_e
8*a*	0.116	0.043	0.136	0.050	0.146	0.023
	0.116	0.043	0.150	0.050	0.146	0.023
8*b*	0.116	0.043	0.136	0.050	0.210	0.023
	0.116	0.043	0.170	0.050	0.096	0.023
9*a*	0.093	0.043	0.136	0.000	0.036	0.050
	0.093	0.043	0.050	0.000	0.150	0.050
9*b*	0.093	0.043	0.136	0.000	0.136	0.050
	0.093	0.043	0.150	0.000	0.150	0.050
9*c*	0.093	0.043	0.136	0.000	0.236	0.050
	0.093	0.043	0.250	0.000	0.150	0.050
10*a*	0.116	0.043	0.136	0.050	0.100	0.023
	0.116	0.043	0.100	0.050	0.096	0.023
10*b*	0.193	0.043	0.136	0.100	0.150	0.050
	0.193	0.043	0.150	0.100	0.150	0.050
10*c*	0.293	0.043	0.136	0.150	0.250	0.100
	0.293	0.043	0.150	0.150	0.250	0.100
11*a*	0.116	0.043	0.186	0.023	0.186	0.050
	0.116	0.043	0.186	0.023	0.186	0.050
11*b*	0.216	0.093	0.286	0.023	0.286	0.100
	0.216	0.093	0.286	0.023	0.286	0.100
11*c*	0.296	0.093	0.286	0.123	0.286	0.080
	0.296	0.093	0.286	0.123	0.286	0.080

NOTES

The values of delay in the column headed τ_w include the delay of the tie line and the added artificial delay.

Circuit 3*b* arranged with relays h_e and h_w to short the echo suppressor without shorting the transmission path.

nation. Observations were also made from time to time on the tie line without added delay and with a single echo suppressor of special design. Since this condition was not subject to lockout, these observations may be used to give an indication of the seasonal effects, and to correct data obtained from the different groups of tests. These data are designated by the letter *n*.

The data recorded consist of the duration of each call, the number of lockouts per call, the total locked out time per call, and the number of repetitions per call. Calls having a duration less than 100 seconds are not included in the data. Table II gives the number of calls observed, the mean duration of the calls in seconds, the number of lockouts per

100 seconds ($L/100$), the per cent of time locked out, or locked out time in seconds per 100 seconds ($LT/100$), the number of repetitions per 100 seconds ($R/100$), and the number of repetitions per 100 seconds cor-

TABLE II

Condition	Number of Calls	Mean Duration Seconds	$\frac{L}{100}$	$\frac{LT}{100}$	$\frac{R}{100}$	$\frac{R'}{100}$
1*a*	275	451	2.61	1.13	0.40	0.40
1*b*	275	437	2.63	1.40	0.44	0.44
1*c*	275	456	3.68	2.34	0.53	0.53
1*n*	275	401			0.36	
2*a*	200	439	2.62	1.03	0.36	0.36
2*b*	200	410	2.21	1.19	0.47	0.47
2*c*	200	393	3.86	1.54	0.45	0.45
3*a*	300	424	3.61	2.34	0.51	0.51
3*b*	300	397	5.90	1.85		
4*a*	275	457	3.26	1.63	0.53	0.51
4*b*	275	413	3.34	1.76	0.55	0.53
4*c*	275	432	3.13	2.02	0.55	0.53
4*n*	75	396			0.38	
5*a*	275	436	3.99	1.99	0.60	0.53
5*b*	275	425	4.03	2.37	0.56	0.49
5*n*	250	392			0.43	
6*a*	50	391	8.34	4.26	0.72	0.67
6*n*	125	390			0.41	
7*a*	200	410	3.00	0.79	0.52	0.41
7*b*	275	387	4.26	1.23	0.55	0.44
7*n*	75	395			0.47	
8*a*	150	387	2.51	0.74	0.49	0.36
8*b*	300	401	4.64	1.30	0.53	0.40
8*n*	100	393			0.49	
9*a*	150	413	0.83	0.02	0.43	0.36
9*b*	150	403	2.59	0.17	0.42	0.35
9*c*	150	380	5.42	1.00	0.44	0.37
9*n*	175	399			0.43	
10*a*	300	401	3.84	0.91	0.46	0.44
10*b*	300	405	4.46	1.78	0.44	0.42
10*c*	300	420	5.28	2.66	0.54	0.52
10*n*	175	439			0.38	
11*a*	300	413	1.64	0.64	0.42	0.37
11*b*	300	408	0.94	0.56	0.38	0.33
11*c*	300	430	2.99	2.28	0.56	0.51
11*n*	300	396			0.41	

rected for seasonal variations ($R'/100$). These rates are obtained by dividing the total number of occurrences, or locked out time by the total duration for each test condition.

Effect of Lockout on Repetition Rate

It would be reasonable to expect that the repetition rate would depend not only on the lockout rate, but also on the duration and type of lockout. Considering the data as a whole there does not appear to be

any definite relation between the lockout rate and repetition rate, although in most cases an increase in lockout rate results in an increase in repetition rate. If we exclude from consideration those cases in which the lockouts are of very short duration and in which releasing lockouts occur, the data indicate a somewhat closer dependence of repetition rate on lockout rate. This suggests that the increase in repetition rate caused by lockouts may be proportional to the duration of lockouts and to their frequency of occurrence, or to the per cent of time which is locked out. Fig. 3, which shows the repetition rates,

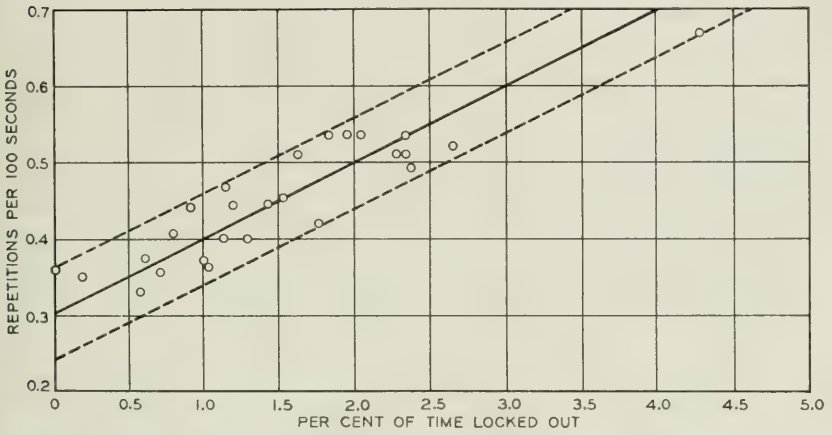


Fig. 3—Observed variation of corrected repetition rate with per cent of time locked out.

corrected for seasonal variations, plotted against the per cent of time locked out, indicates a reasonable agreement with this assumption. The correction is applied by subtracting from the observed repetition rate the difference between the observed repetition rate for the appropriate reference or *n* condition and a rate of 0.36, arbitrarily chosen as equal to the lowest repetition rate observed on any of the *n* conditions. All of the data are included in this figure. The dashed lines are drawn to include all the data and have a slope estimated as average from considering the data in individual groups. The variability in the data may in part be attributed to the variation in the distribution of lockout durations, since if two distributions have the same mean value but different spreads, the lockouts comprising the distribution which includes a greater number of long lockouts might be expected to have a greater effect upon the repetition rate. With due allowance for the variability of the data, Fig. 3 indicates that the repetition rate increases proportionally with the per cent of time locked out except

possibly for values less than 0.6 per cent. The slopes of the boundary lines are such as to show about 0.1 increase in repetition rate with each 1 per cent of locked out time, and this relation appears to hold for releasing as well as lasting lockouts, and for lockouts which may be caused by relay operations by noise.

Certain qualifications are necessary in considering the significance of this result. The indicated increase in repetition rate may be partly due to other causes than lockout, as for example the effects introduced by the delay of the circuit, or by the relay hangover, during changes in the direction of speech transmission which are not accompanied by lockout. The net effect of these causes increases with circuit changes which increase the per cent of time locked out. Consequently, the latter may be taken as a criterion of the total effect, even though the contribution of the former to the repetition rate may be appreciable.

No general significance can be attached to the absolute values of the repetition rates observed in these tests since it is well known that repetition rates will differ for identical circuit conditions used with different terminal conditions and by different classes of telephone subscribers. These observed rates are significant only for comparing the relative performance of circuits under the particular conditions of use pertaining to these tests.

The significance of the results obtained depends upon the assumption that a change in lockout which causes an increase in repetition rate is an undesirable change and the transmission performance is thereby degraded. In the case of certain circuit changes which introduce changes in intelligibility the resulting changes in repetition rate can be used to determine effective transmission ratings,[5] expressed in db, of the circuits under consideration. A corresponding procedure might be applied to express the observed changes in repetition rate due to lockout in terms of db, but in the absence of data to establish the equivalence of the ratings for different types of degradation, it has not seemed advisable to do so.

Locked Out Time in Terms of Circuit Constants

Since these tests indicate that the repetition rate is proportional to the per cent of time locked out we can limit our consideration to the latter as a suitable criterion for measuring the relative merit of circuits equipped with two echo suppressors. To determine the per cent of time locked out we can measure it directly, as has been done in these tests, or it can be calculated in terms of the circuit constants by deter-

[5] "Scientific Research Applied to the Telephone Transmitter and Receiver," Edwin H. Colpitts, *Bell System Technical Journal*, July 1937.

mining the average number of lockouts per hundred seconds and the average duration of lockouts in terms of the circuit constants and obtaining the per cent of locked out time as the product of these two quantities.

The average, or expected number of lockouts per hundred seconds can be approximately determined from the circuit constants and the distributions of response and resumption times. It is shown in the appendix that, subject to certain assumptions, the probability of lockout following a pause is given by

$$P = \int\int p_1(x)\, p_2(y)\, dx\, dy, \tag{1}$$

in which $p_1(x)\, dx$ and $p_2(y)\, dy$ are the probabilities that, following a pause, the resumption time will be between x and $x + dx$ and the response time will be between y and $y + dy$. As suitable approximations to these probabilities we may take the observed distributions of resumption and response times. Mr. Norwine and Mr. Murphy, in their accompanying paper,[3] give distributions of resumption and response times which are shown in Fig. 4. These distributions are expressed in terms of the total number of resumptions, or responses and consequently the data are an approximation to the conditional probability that if a resumption or response has occurred, the resumption or response time will be between t and $t + dt$. The use of these data will

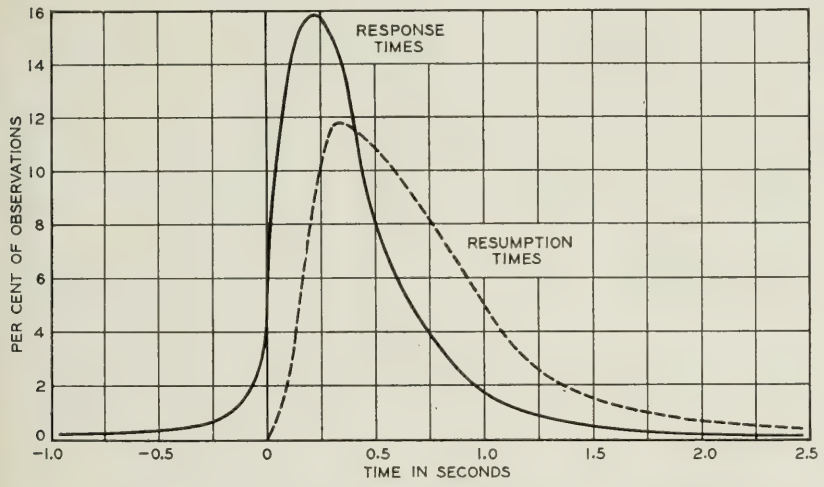


Fig. 4—Observed distribution of resumption and response times.

[3] Loc. cit.

therefore result in calculated values of the probability of lockout which are proportional to the desired probability, and if the value of the integral calculated from their data is p, then

$$P = kp, \tag{2}$$

where k is a constant of proportionality which depends on the average number of pauses occurring, and which can be determined by comparing observed and calculated results.

The observed lockouts per hundred seconds plotted against the calculated probability of lockout for each circuit condition are shown in Fig. 5 for lasting lockouts and in Fig. 6 for releasing lockouts. The

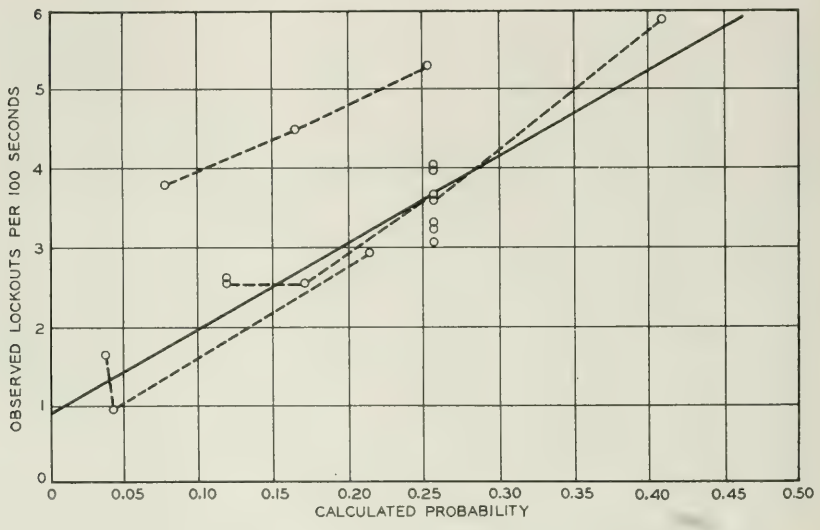


Fig. 5—Observed lasting lockouts vs. calculated probability.

data are separated in this way, since the factor of proportionality between observed and calculated results is found to be different for the two cases. This is probably due to the fact that the method of determining response and resumption times was such that some of the negative and shorter positive response times could not be detected. An increase in the number of these response times would result in an increased probability of releasing lockouts, which would tend to bring the two sets of data into agreement. Greater accuracy might be obtained if the distribution of response times were to be more accurately determined, but the present data are sufficient for approximate calculations.

In both figures the data obtained in a single group of tests are connected by dotted lines. The solid lines are the best estimates to repre-

sent the two complete sets of data. In the case of Fig. 5 the solid line was determined by the method of least squares, omitting the data of group 10. This omission appears to be justified since these data are consistent among themselves and yield a factor of proportionality

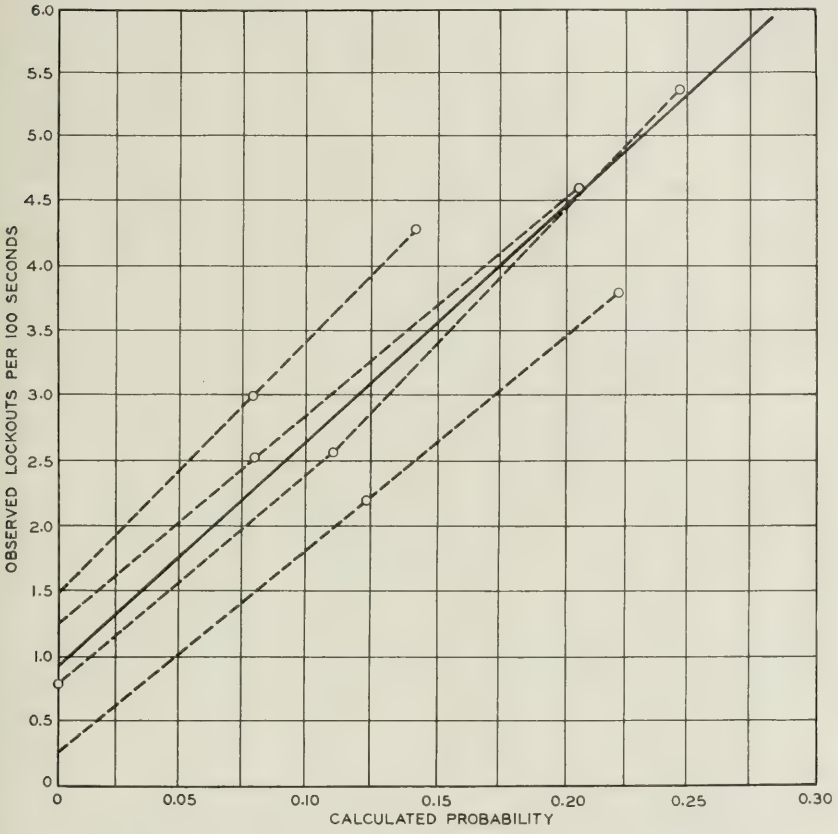


Fig. 6—Observed releasing lockouts vs. calculated probability.

which is consistent with the rest of the data, and since it is known that many uncontrolled factors may influence the results of one particular group of tests. In the case of Fig. 6 the solid line was obtained by averaging the slopes and constant terms of the individual dotted lines.

Both sets of data indicate that about 0.9 lockout per hundred seconds occurs when the calculated probability of occurrence is zero. This is undoubtedly due to non-synchronous action of the suppressors, caused by slight variations in sensitivity, changes in effective hangover caused by changes in sensitivity and by occasional relay chatter. This con-

clusion is confirmed by the tests made with no delay between the suppressors, in which lockout is obviously impossible with synchronous action, but in which lockouts were actually obtained in amount consistent with the rest of the data.

Figures 5 and 6 show that the number of lasting and releasing lockouts can be calculated from the circuit constants and the distributions of response and resumption times. Approximations which are suffi-

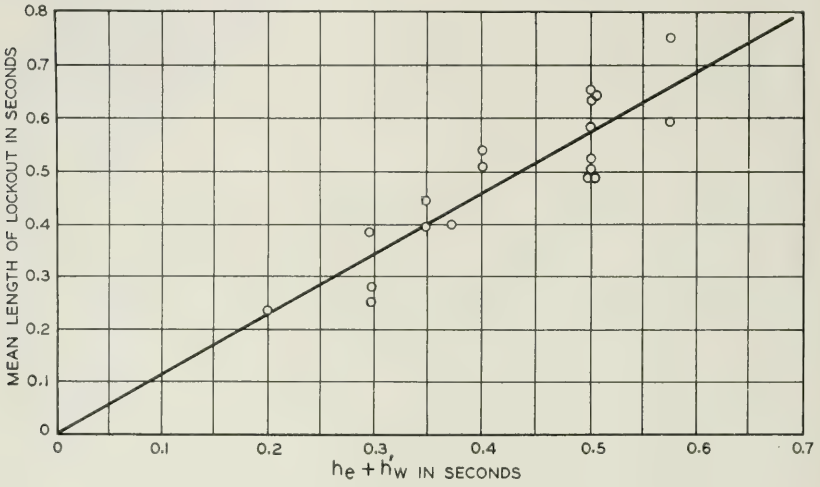


Fig. 7—Observed length of lockout as a function of relay hangovers.

cient for practical purposes for calculating the number of lasting and releasing lockouts are respectively

$$L_l = 0.9 + 11.0\,p, \tag{3}$$

$$L_r = 0.9 + 17.7\,p. \tag{4}$$

The duration of a lockout is obviously dependent upon the way in which the subscribers talk and upon the hangovers of relays h_w and h_e'. Lockouts of several seconds duration have frequently been observed but most frequently the duration of lockouts appears to be short and determined primarily by the relay hangovers. Figure 7 shows the mean duration of lasting lockouts plotted as a function of the sum of the relay hangovers, $h_e + h_w'$. The straight line is the least square representation of the data, which is

$$D_l = 0.002 + 1.16\,(h_e + h_w'). \tag{5}$$

The constant term in this equation can be neglected for approximate calculations.

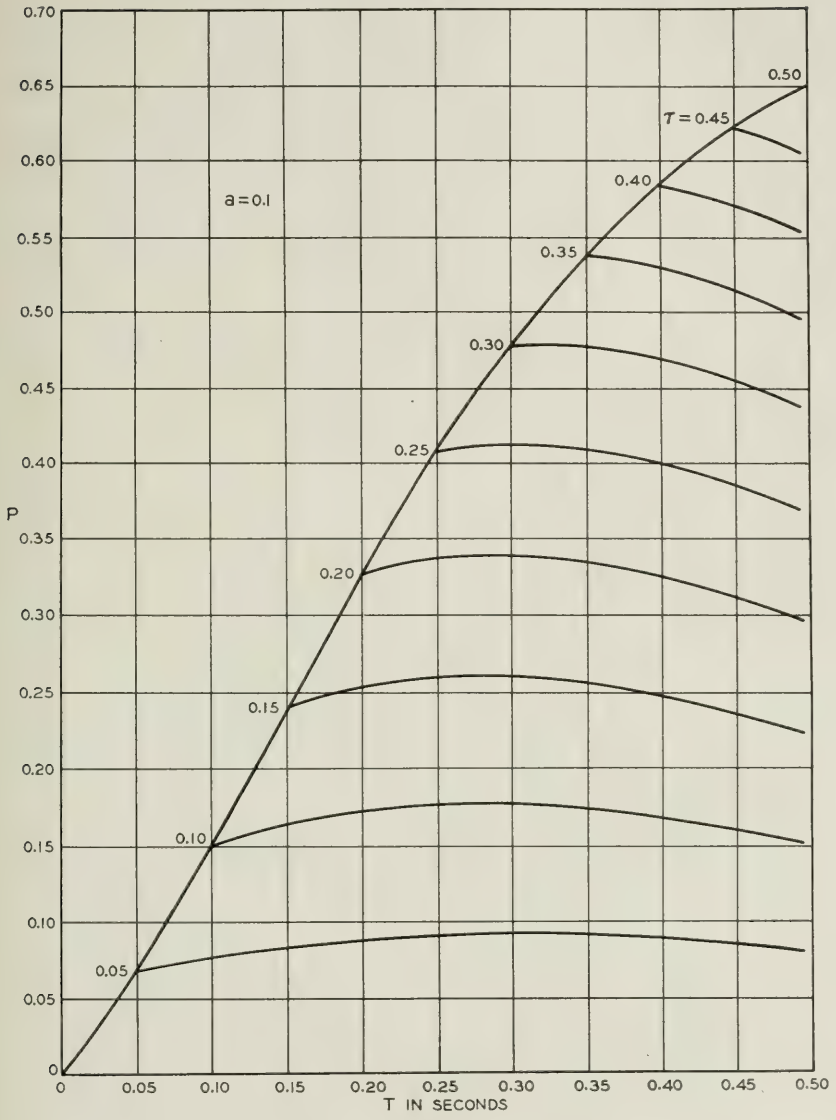


Fig. 8—Calculated probability of lockout as a function of total circuit delay.

Since the duration of lasting and releasing lockouts is not the same, releasing lockouts being of short duration, and since the two were not separately observed, the data are insufficient to determine the duration of releasing lockouts directly. However, tentative calculations indicate that approximate results can be obtained by assuming that the

mean duration of releasing lockouts is about one-quarter that of lasting lockouts or

$$D_r = 0.29\ (h_e + h_w'). \tag{6}$$

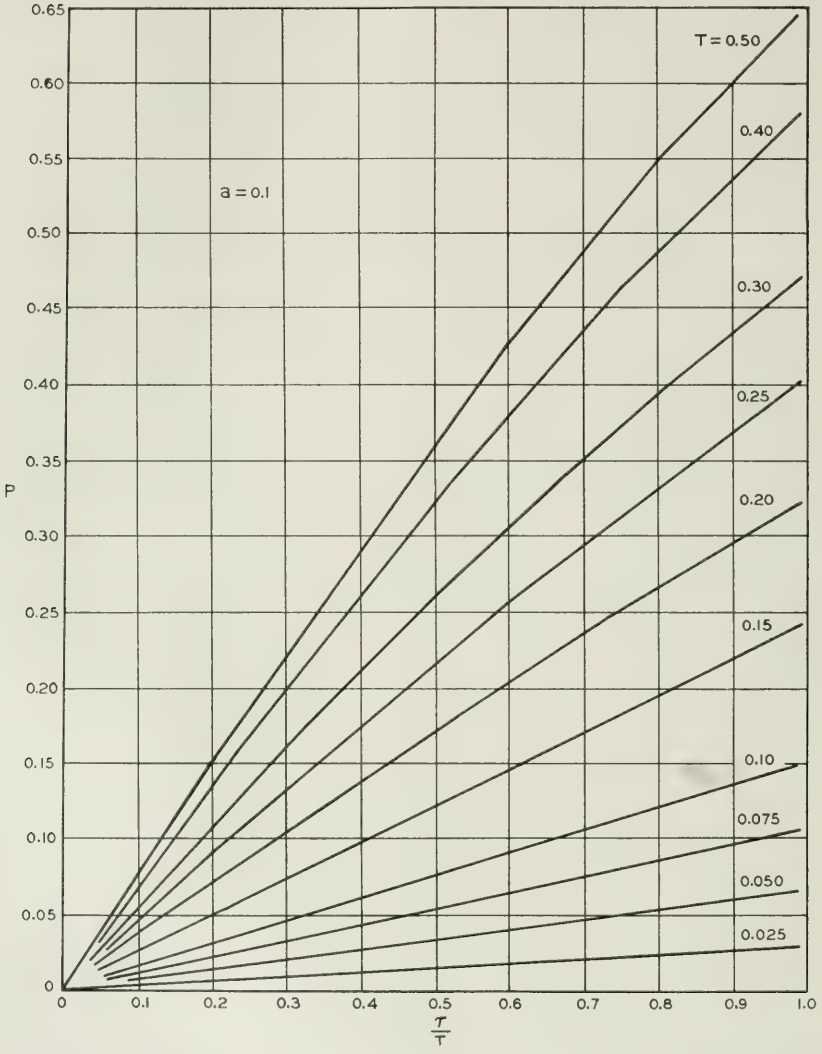


Fig. 9—Calculated probability of lockout as a function of the ratio of the delay between the suppressors to the total circuit delay.

With the above relations between probability of lockout and circuit constants, number of lockouts and probability of lockout, and mean duration of lockouts and relay hangovers, it is possible to determine the

per cent of time locked out from the circuit constants. Since this is proportional to the repetition rate, a measure of relative circuit performance is obtained in terms of the circuit constants.

As an example of such calculations let us assume that $\tau_w = \tau_w' = \tau_e = \tau_e'$, $\tau = \tau'$ and $T = T'$, and $h_w = h_e' = 2\tau_w + a$. Then the constants of integration determined in the appendix become,

$$\begin{aligned} a &= a, \\ b &= a + T - \tau, \\ c &= a + T + \tau, \end{aligned}$$

and the probability of a lasting lockout is proportional to

$$P = \int_{-\infty}^{a} p_2(y)dy \int_{a+T-\tau}^{a+T+\tau} p_1(x)dx + \int_{a}^{\infty} p_2(y)dy \int_{y+T-\tau}^{y+T+\tau} p_1(x)dx. \quad (7)$$

Values of this probability for $a = 0.1$ are shown in Fig. 8 as a function of the transmission time T with τ, the delay between the suppressors as a parameter and in Fig. 9 as a function τ/T with T as a parameter. The curves are not extended beyond $T = 0.5$ since smaller values are thought to cover the range of practical interest. Furthermore, for large values of T there is some evidence that the effect of the transmission time would be noticed by the subscribers with a consequent change in the distributions of resumption and response times.

These curves indicate that for a constant value of τ, the delay between the echo suppressors, there is little change in the probability of lockout as the total delay of the circuit T is increased, and for a constant value of T the probability of lockout is approximately proportional to τ.

To continue with a more specific example, let us consider a telephone connection consisting of two four-wire circuits each equipped with an echo suppressor in the center of the circuit as shown in Fig. 10. In the notation of Fig. 10 the relay hangovers are each equal to $\tau + 0.100$ and the constants of integration are

$$\begin{aligned} a &= 0.100, \\ b &= 0.100 + \tau, \\ c &= 0.100 + 3\tau. \end{aligned}$$

Since the two circuits are assumed equal, only lasting lockouts are theoretically possible and the curves of Fig. 8 may then be used to determine p in terms of τ as defined by equation (7), which in turn may be used to determine the expected number of lockouts from equation (3). The mean duration of lockout is obtained by inserting the value

of the relay hangover in equation (5) giving

$$D = 0.234 + 2.32\,\tau.$$

The product of D with the expected number of lockouts per hundred seconds is then equal to the per cent of time locked out, which is shown in Fig. 10 as a function of τ. By using the relation shown in Fig. 3 a

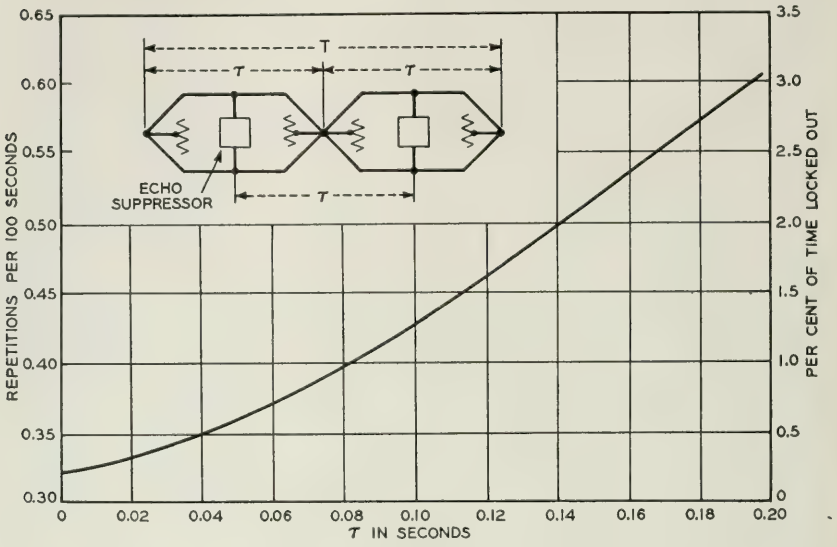


Fig. 10—Calculated per cent of time locked out, and repetition rate for the indicated circuit conditions.

second scale is shown in Fig. 10 to give the relation between the repetition rate and the delay between the suppressors. This curve shows that the repetition rate increases with the delay between the suppressors, at a gradually increasing rate up to a delay of about 0.09 seconds, beyond which the impairment increases linearly with the delay.

Summary

It has been shown that two types of lockout, lasting and releasing lockouts, may occur in telephone connections involving two echo suppressors, and the manner of their occurrence has been discussed.

The results of an experimental investigation show that the occurrence of lockouts causes an increase in repetition rate, which is approximately proportional to the per cent of time locked out.

There has been presented a theoretical method for calculating the expected number of lockouts in terms of the circuit constants which de-

pends upon the characteristic time intervals in conversational speech. The values which have been calculated with experimentally determined constants are shown to agree with the observed values.

The average duration of lockouts has been found to be proportional to the hangovers of the relays effective in lockout.

Since the per cent of time locked out is equal to the product of the average number of lockouts per hundred seconds and the average duration of lockouts, it may be determined in terms of the circuit constants, and used as one of the criteria of the relative performance of the circuits under consideration.

Specific examples of such calculations have been used to illustrate the relations between the expected number of lockouts and the circuit constants, and between the repetition rate and the constants of a particular circuit configuration.

Subject to certain restrictions on the relations between the circuit constants, it appears that the number of lockouts and the resulting increase in repetition rate are approximately proportional to the delay included between the echo suppressors.

In conclusion I wish to express my appreciation to my associates who have contributed to this study; in particular to Dr. G. R. Stibitz who first developed the theoretical approach to the problem, to Mr. W. R. Bennett and Mr. B. D. Holbrook who have contributed to the extension of this approach, and to Mr. A. C. Norwine and Mr. O. J. Murphy who obtained the distribution functions used in the calculations and conducted the experimental work.

Appendix

To assist in formulating an expression for the probability of lockout a number of simplifying assumptions have been made, as follows:

Pauses in speech are sufficiently separated to be considered as independent events, or in other words the sequence of events occurring at one pause have no effect upon those occurring at another.

Following a pause each speaker can start speaking only once and only one of three events can occur.

1. The original speaker regains control of the circuit.
2. The other speaker obtains control of the circuit.
3. Lockout occurs.

Resumption and response times are independent.

The distributions of response and resumption times are independent of the delay of the circuit and the disposition of the echo suppressor.

The operate times of the suppressors are sufficiently small to be neglected.

Let $p_1(t)dt$ be the probability that the speaker in control of the circuit will resume speaking in the interval t to $t + dt$ after pausing and let $p_2(t)dt$ be the probability that the speaker not in control of the circuit will start speaking in the interval t to $t + dt$ after hearing the other speaker pause. In the latter case t may be negative. Then the probability that, following a pause, a resumption will occur in the interval x to $x + dx$ and a response will occur in the interval y to $y + dy$ is given by

$$p_1(x)\ p_2(y)\ dx\ dy,$$

and the probability of lockout following a pause is given by

$$P = \int\int p_1(x)\ p_2(y)\ dx\ dy,$$

in which the integration is to be performed over the region in the xy plane which contains those values of x and y for which lockout occurs.

Assuming that either subscriber is equally likely to have control of the circuit at any instant, the probability of a lockout following a pause by either party is given by the average of the probabilities for the two parties.

In determining the limits of integration there are three cases to be considered. Assuming a pause by E

$$\begin{aligned} &\text{I.} && h_e < h_w + \tau, \\ &\text{II.} && h_w + \tau < h_e < h_w + \tau + \tau', \\ &\text{III.} && h_w + \tau + \tau' < h_e. \end{aligned}$$

In case I only lasting lockouts can occur while in cases II and III both lasting and releasing lockouts can occur. Case I will be used to illustrate a method of determining the limits of integration which can also be applied to cases II and III for which the results will be stated without proof.

In Fig. 11 which is based on the circuit of Fig. 1, time is represented horizontally and distances vertically, upward from the central line, which represents the east end of the circuit, for transmission from W to E and downward for transmission from E to W. Consider a pause by E occurring at $x = 0$ at A. The line $ABDG$ represents the transmission of this pause to W. The point H obtained by projecting G to the top line determines the point $y = 0$. The points C and F represent the instants at which h_e and h_w release. If E resumes in the interval AI, the resumption will arrive at the input of h_w before h_w has released as determined by the point F, and E retains control of the circuit. If, on the other hand W responds at any time prior to

J the response will be blocked by h_w until the time represented by K and it will then be transmitted to the end of the circuit as shown by the line $JKLM$. If now E resumes at any time after P the resumption

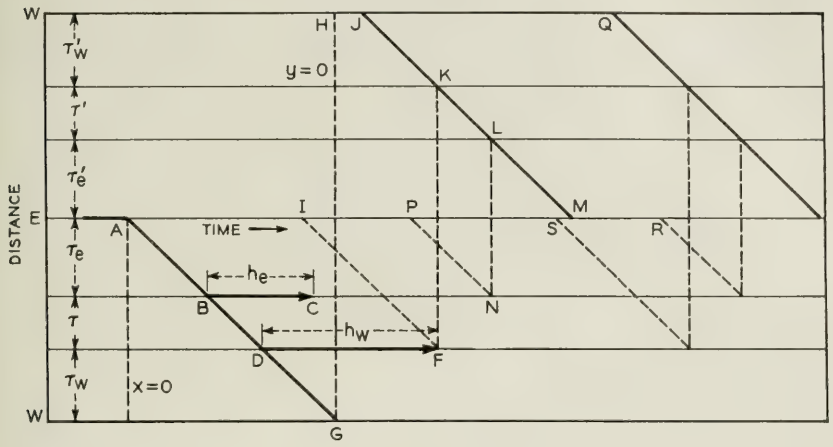


Fig. 11—Time relations in four-wire circuit for determination of limits of integration.

will be blocked by h_e' since W will have obtained control of the circuit. A resumption in the interval IP will result in lockout since W will control h_w' and E will control h_e. If W responds at some time after J, say at Q, a similar argument can be used to show that E must resume in the interval SR to cause lockout. If we let

$$\begin{aligned} IIJ &= a, \\ AI &= b, \\ AP &= c, \end{aligned}$$

it can be shown that

$$\begin{aligned} AS &= y - a + b, \\ AR &= y - a + c, \end{aligned}$$

and therefore the region of integration is defined by

$$\begin{aligned} b < x < c, \qquad -\infty < y < a, \\ y - a + b < x < y - a + c, \qquad a < y < \infty, \end{aligned}$$

in which a is the time interval the speaker not in control must wait after hearing the other speaker pause in order to enable the response to get through the circuit; b is the time interval after the speaker in control pauses, during which he can gain control by resuming, regardless of what the other speaker does; c is the time interval the speaker in control must pause in order to make it possible for the other speaker to get a response through the circuit.

These constants have the values

$$\begin{aligned} a &= h_w - (\tau_w + \tau_w'), \\ b &= h_w, \\ c &= h_w + (\tau + \tau'). \end{aligned}$$

Case II. By the same method the regions of integration are determined to be, for lasting lockouts

$$\begin{aligned} &y < a, \qquad b < x < c, \\ &y < a, \qquad y - a + b < x < y - a + c, \end{aligned}$$

for releasing lockouts, blocked by h_w,

$$y < a, \qquad \beta < x < \infty,$$

for releasing lockouts without blocking,

$$a < y < \alpha, \qquad \beta < x < \infty,$$

in which α and β are defined by

$$\begin{aligned} \alpha &= h_e - \tau - (\tau_w + \tau_w'), \\ \beta &= h_e. \end{aligned}$$

Case III. As before we obtain the regions, for lasting lockouts

$$\begin{aligned} &y < a, && b < x < \beta, \\ &a < y < \gamma, && y - a + b < x < \beta, \\ &\gamma < y < \infty, && y - a + b < x < y - a + c, \end{aligned}$$

for releasing lockouts blocked by h_w and h_e,

$$y < a, \qquad \beta < x < \infty,$$

for releasing lockouts blocked by h_e,

$$a < y < \gamma, \qquad \beta < x < \infty,$$

for releasing lockouts without blocking,

$$\gamma < y < \alpha, \qquad \beta < x < \infty,$$

in which α and β have the values given above and

$$\gamma = h_e - (\tau + \tau') - (\tau + \tau_w').$$

Characteristic Time Intervals in Telephonic Conversation

By A. C. NORWINE and O. J. MURPHY

Two-way conversation is arbitrarily defined in terms of vocal intervals and the pauses between them. These quantities, as determined by the presence or absence of speech energy, have been measured from continuous oscillograms of calls on a New York-Chicago telephone circuit used for Bell System business, and the results of statistical analyses of these data are presented.

Introduction

THE time pattern of a conversation may be described in terms of the periods during which speech energy is issuing from the lips of each talker, the pauses with which each intersperses his speech, and the periods after the termination of a talker's speech during which the listener prepares to reply. On a telephone circuit this can be determined by the presence or absence of speech energy within the circuit, measured by an appropriate recording instrument. It is with observations of this type and the measurement of time intervals in conversation obtained in this manner with which the present paper is concerned.

It will be well to keep in mind that the fundamental basis of these measurements is the presence or absence of speech energy. Many of the pauses recorded in this study are of the type which are known to occur within sentences, phrases, or even within words. Some of these are insufficient in duration to interrupt the continuity of the flow of speech, and some are too short to be noticed by a listener. The intervals as defined in this paper probably do not, therefore, exactly correspond to those which would be observed by a person listening to the conversations.

The study and measurement of these intervals were originally undertaken to furnish information needed in the application of probability theory to the occurrence of lockouts on toll telephone circuits equipped with tandem voice-operated devices. This problem is treated in a companion paper by Mr. A. W. Horton, Jr.* Since that time, however, parts of the data have been used in various other technical applications, and it is therefore thought that the results of the study may have some general interest.

* "The Occurrence and Effect of Lockout in Telephone Connections Involving Two Echo Suppressors," Arthur W. Horton, Jr., this issue of the *Bell System Technical Journal*.

Nature of the Problem

In the simplest case of conversational interchange each party speaks for a short time, pauses, and the other party replies. The time intervals are then simply the lengths of time each party speaks and the lengths of the pauses between speeches. The period during which there is speech may be called a talk-spurt, and the length of the pause may be called the response-time. These two quantities would then suffice to describe this simple type of interchange.

In many instances, however, the process is not so orderly; for example one speaker may pause and then resume speaking, or the listener may begin to reply without waiting for the end of the talker's speech. The possible, and indeed frequently encountered, variations of the simple cycle of which the preceding examples constitute only a fraction make it necessary to carefully define and delimit the elements into which a conversation may be resolved. It is believed that any telephonic conversation between two persons can be completely described in terms of the presence or absence of energy by the following time elements:

> A *talkspurt* is speech by one party, including his pauses, which is preceded and followed, with or without intervening pauses, by speech from the other party perceptible to the one producing the talkspurt. Obvious exceptions to this definition are the initial and final talkspurts in a conversation. There may be simultaneous talkspurts by the two talkers; if one party is speaking and at the same time hears speech from the other *double talking* is said to occur.
>
> *Resumption time* is the length of the pause intervening between two periods of speech within a talkspurt.
>
> *Response time* is the length of the interval between the beginning of a pause as heard by the listener and the beginning of his reply. It may be positive or negative. The pause to which reference is made ordinarily occurs at the end of a talkspurt but may be a pause followed by a resumption of speech by the first talker.

In the terms of these definitions a telephone subscriber "hears" or "perceives" when voice currents flow in his receiver; a possible lack of attention or other failure to appreciate what is "heard" is not considered. Likewise, it should be stressed that pauses are not confined to the intervals between words or sentences but may occur within words; in the measurements described herein they are determined solely by the absence of voice energy on the circuit.

In addition to these natural conversational elements there is a fourth

item which is sometimes imposed by the configuration of toll circuits, namely *lockout*.[1] Lockouts did occur on the circuit carrying the conversations which provided the data for this study, due to the special circuit arrangement employed at the time, but their occurrence will not be treated in this paper.

The specific information desired was the probability that a conversational element would have any given duration t. The true probability in each case can be approximated, except for a scale factor, by a distribution curve of experimental data depicting the fractions of the total number of observations of the quantity which lie within each of a regular progression of time cells of a chosen width. It is in this form that the data will be presented.

Data Source

Some preliminary data were obtained from observations on local inter-office calls between various members of the staff of Bell Telephone Laboratories. Members of the non-technical groups were included among the talkers, some of whom were women. Equipment added to the telephone circuit for the purpose of recording was held at a minimum and its presence and action were not noticed by any of the talkers. The recording means were less elaborate than those employed in the later investigation, and the principal interest in the preliminary test lies in the fact that the results have shown themselves in good agreement with those obtained later with different talkers and widely different circuit conditions.

The conversations which provided the material for the main part of this study were those which took place between male talkers on a circuit used as a tie line by the Western Electric Company and running from the company's Hawthorne plant at Chicago, Illinois, to the New York office. This is a circuit which is used wholly for transaction of company business, and most of the users hold at least minor executive positions in their organizations and have a background of scientific or formal business training. It is recognized by the authors that this somewhat restricted class of talkers may not be representative of telephone users in general. It is also recognized that the manner of telephonic conversation may be different for long distance and local calls. However, analyses of data from other tests which have been made on the tie line, involving a wide range of delays and types of voice operated devices, have indicated that the talkers endeavored to converse in the same manner regardless of the circuit configuration.

[1] A *lockout* is the simultaneous blocking, by voice-operated devices, of both directions of transmission of a two-way communication system. For a discussion of this, and other possible definitions, see the companion paper by Mr. A. W. Horton, Jr.

The Western Electric circuit is a four-wire 19-gauge H-44 circuit [2] 817 miles long, with a 1000-cycle time of transmission of 0.043 second. Normally there is an echo suppressor at Pittsburgh, which is approximately at the midpoint of the circuit. In connection with other tests which were going on at the time, the normal echo suppressor was removed, and the circuit was looped via Bell Telephone Laboratories where artificial (acoustic) delay circuits and two echo suppressors were inserted to simulate two tandem circuits, each with an echo suppressor at its midpoint. The extra equipment introduced no additional attenuation or frequency discrimination.

Recording Methods and Mechanisms

All of the recording was done by mechanical means controlled by engineers who observed the progress of the conversations. The preliminary data on local calls were obtained with the aid of an inked-roller paper-tape recorder of a type formerly used in telegraph studies. This machine had only two recording traces and was not adapted to run at high speed, thus limiting the amount and accuracy of the information obtainable. In view of the limited amount of data and its relative lack of precision compared to the main body of data it does not appear profitable to enter into a further description of the early recording means.

The principal part of the improved recording mechanism was a six-string rapid-record oscillograph of the type already described in the Bell System Technical Journal.[3] The several strings of this machine were energized by speech power from the two talkers and by energy from an oscillator under control of the echo-suppressor relays. This arrangement is indicated in Fig. 1. The machine was started at the beginning of each call to be observed and ran continuously at a speed of about 20 feet of recording paper per minute. This resulted in a complete pictorial record of the conversational interchanges. These records are well adapted to measurement of the essential time relations, but do not lend themselves to reproduction of the original conversations. Operation of the echo-suppressor relays is also shown on the oscillograms, but analysis of this information is outside the scope of this paper.

To facilitate inspection of the speech traces the voice energy from the circuit was routed through quick-acting automatic volume controls which permitted the weak beginnings and endings of words to be

[2] The designation H-44 means 44 millihenry loading coils spaced 6000 feet apart.

[3] "An Oscillograph for Ten Thousand Cycles," A. M. Curtis, Vol. XII, No. 1, pp. 76–90.

observed without producing excessive amplitudes of the recording strings during the strong parts of speech. These devices, as used, distorted the recorded envelope of the speech sounds considerably, but assisted materially in showing just where speech traces started and stopped. Noise occasioned little difficulty in this study. When observable, the traces of the noise were so characteristic that little confusion with speech waves resulted.

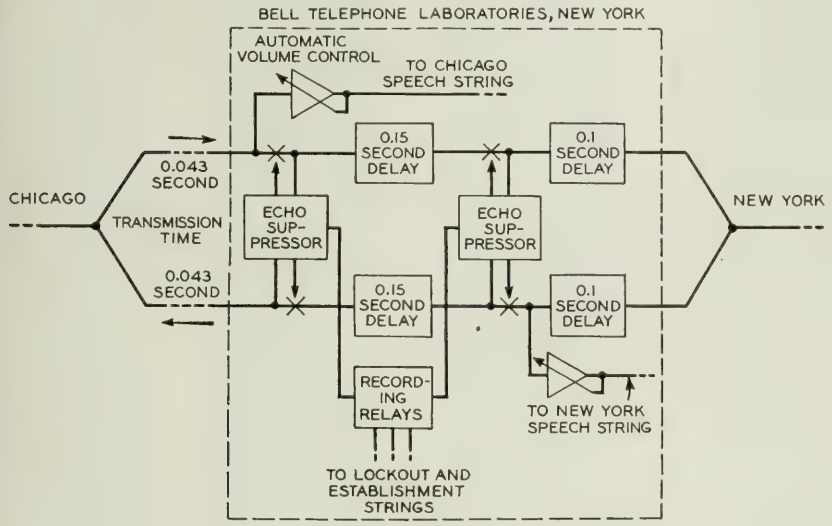


Fig. 1—Connections to New York–Chicago tie line to obtain oscillograms of conversations.

Some observations of response and resumption time were lost when double talking occurred, because echo suppressor operation introduced appreciable loss into one side or the other of the recording circuit.[4] High gain in the automatic volume control and high speech volumes, however, permitted many of such instances to be properly recorded. There is little reason to believe that the instances lost were frequent or that they differed in any particular regard from those recorded; such instances occurred only in the event of double talking involving weak speech by the responding talker.

Instances in which double talking was clearly recorded required a certain amount of arbitrary judgment to decide whether the response was induced by a resumption pause or represented a negative response time. Usually the structure of the conversation was evident, but it is recognized that in some cases there is room for a legitimate difference of opinion.

[4] This loss, while considerable, was not nearly so great as that introduced into the transmission path by the echo suppressor operation.

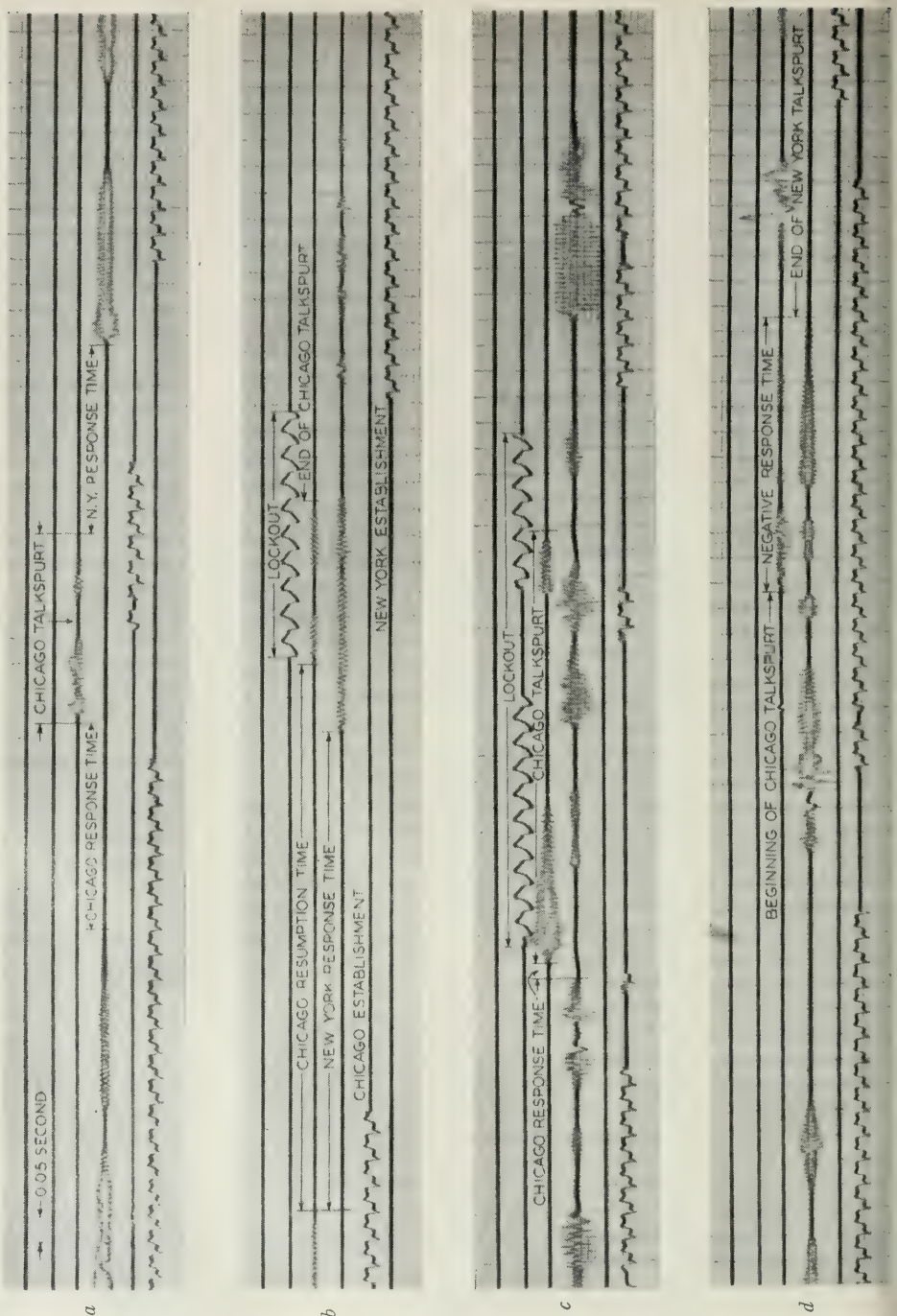


Fig. 2—Typical sections of oscillographic records.

a. Circuit reversals. New York completed talkspurt, heard Chicago's reply, and began another talkspurt.

b. Reply by New York during pause by Chicago caused lockout. New York gained control of the circuit.

c. Short reply by Chicago during pause by New York caused lockout. New York regained control of the circuit.

d. Negative response time. Chicago replied before the completion of New York's talkspurt.

A few samples from the original oscillograms are shown in Fig. 2. The speech energy in each sample is shown on traces 3 and 4 counting from the top down, the upper being from Chicago and the lower from New York. The cyclic waves on traces 2, 5 and 6 indicate respectively lockout, establishment by Chicago and establishment by New York .[5] These waves were obtained from an oscillator which was concurrently used to drive an escapement-type electric clock for measuring the total call duration.

The top oscillogram was selected to show the simplest type of conversational interchange. It will be seen that New York had been talking but had reached the end of his talkspurt as marked on the film. Approximately 0.4 second later Chicago responded, his talkspurt apparently consisting of three syllables, whereupon after a further time of about 0.35 second New York responded and continued talking. The second film was selected to show a less simple type of interchange wherein a long pause within a talkspurt prompted the listener to reply. In this instance the times were such that a lockout resulted. Since the remainder of the talkspurt by the original talker, Chicago, was short and the responding party, New York, continued talking, the circuit was established in New York's direction after the lockout. In the third oscillographic strip Chicago attempted to interrupt, and a short pause by New York permitted lockout to occur; Chicago did not gain control of the circuit. This is an example of concurrent talkspurts, both of which were included in the data. The fourth example was chosen to illustrate a negative response time. In this case Chicago began to reply before the end of New York's talkspurt; no lockout occurred, but the first part of the reply was inaudible to New York due to continued establishment of the circuit in the opposite direction.

It may be noted in Fig. 1 that speech from Chicago was recorded 0.25 second before it was heard by New York and that speech from New York was recorded 0.193 second before it arrived at Chicago. Likewise the beginning of each response did not occur at the time shown on the oscillograms but at a time previous by the delay from the talker's position to that of the recording means. To obtain the response times as previously defined each apparent response time was given an appropriate time correction.

Data Obtained

The more detailed observations were made on fifty-one calls with a total recorded duration of a little over 13,000 seconds. At the recording speed of 20 feet per minute this resulted in about 4400 feet of

[5] An establishment by a talker is said to occur when his speech energy has gained control of all voice operated equipment in his transmission path.

oscillograms. In all cases recording began at the start of the call, but in some instances recording was stopped before the termination of the call due to lack of recording paper in the oscillograph. The oscillograms, ranging in length from 29.6 to 660.8 seconds, represented observations on calls whose mean duration was 430.5 seconds. The speed of recording was such that the time intervals under observation could readily be measured with a precision of ± 0.005 second. The conversational elements were measured with this precision and listed in their order of occurrence for each call. The records for all calls were

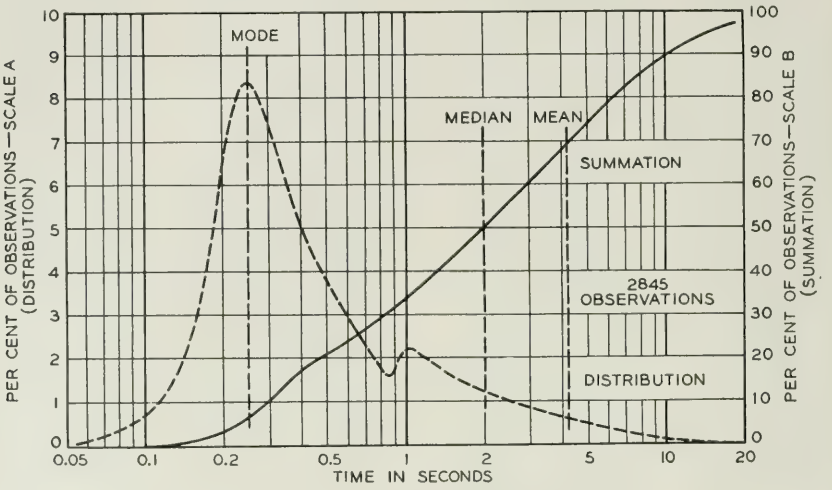


Fig. 3—Lengths of talkspurts.

then consolidated and retabulated in terms of the number of instances of each element whose duration could be included within each of a regular progression of time increments. For all three items of data time cells 0.10 second wide were chosen. The data, when thus cellularized, provided the basis for the construction of histograms from which the time-distribution curves were obtained. These distribution curves and their respective summation curves are given in Figs. 3, 4, and 5. Some of the statistically significant quantities [6] are tabulated on the opposite page. The values are time intervals in seconds.

Since most telephonic speech syllables are shorter than 0.3 second the modal value of 0.25 second for the length of talkspurts makes it clear that monosyllabic replies are by far the most numerous. From

[6] The *mode* is the value which occurs most frequently, i.e., the peak of the distribution curve.
The *median* is that value above and below which equal numbers of observations lie.
The *mean* is the arithmetic average of all the values observed.

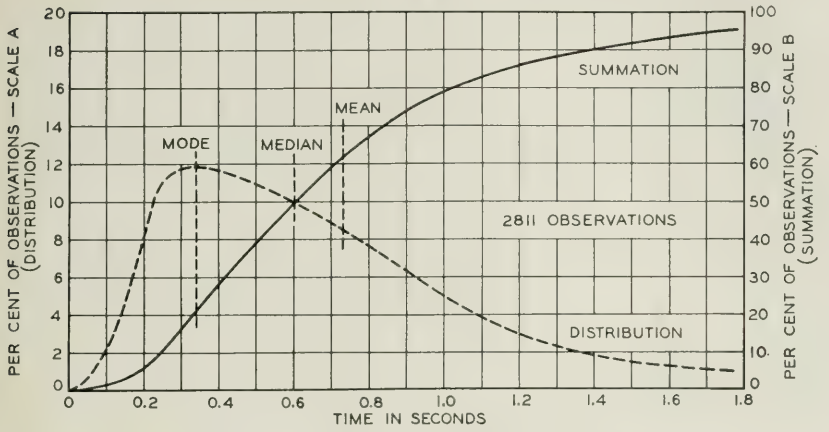


Fig. 4—Lengths of pauses within talkspurts, i.e. *resumption times.*

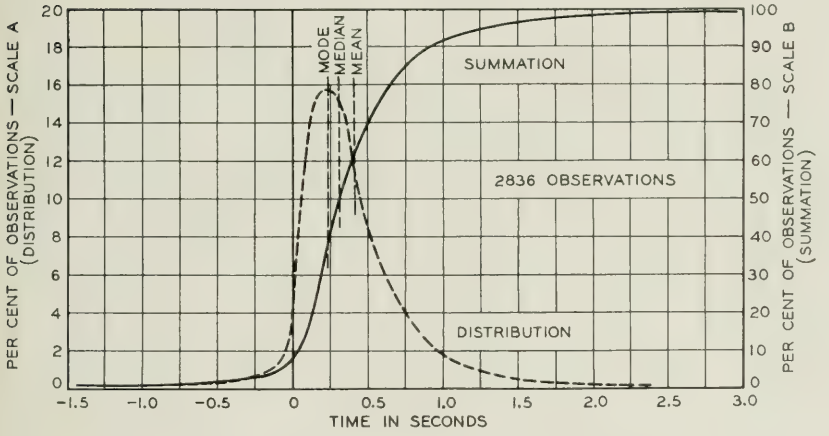


Fig. 5—Lengths of pauses between talkspurts, i.e. *response times.*

	No. of Obs.	Min.	Mode	Median	Mean	Max.
Talkspurts	2845	0.09	0.25	2.00	4.14	143.82
Resumptions	2811	0.05	0.34	0.60	0.73	4.86
Responses	2836	−3.95	0.24	0.32	0.41	5.04

Fig. 3 it may be seen that these, in conjunction with terse replies or questions under one second in duration, constitute about a third of the talkspurts. There were, however, a few very long talkspurts: 27 exceeded 30 seconds, and of these 2 were over 120 seconds long. During the longest talkspurt, which was 143.82 seconds, there were 62 resumptions following silent intervals ranging from 0.34 to 4.04 seconds.

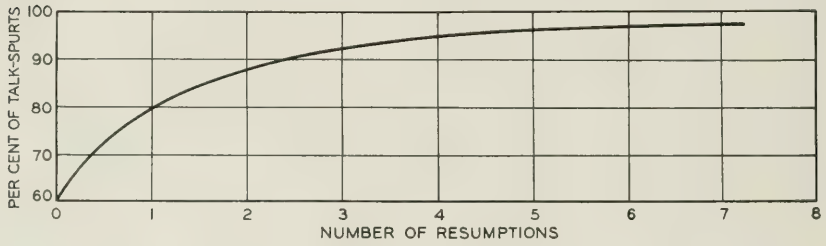


Fig. 6—Percentage of talkspurts containing a number of pauses equal to or less than a given number.

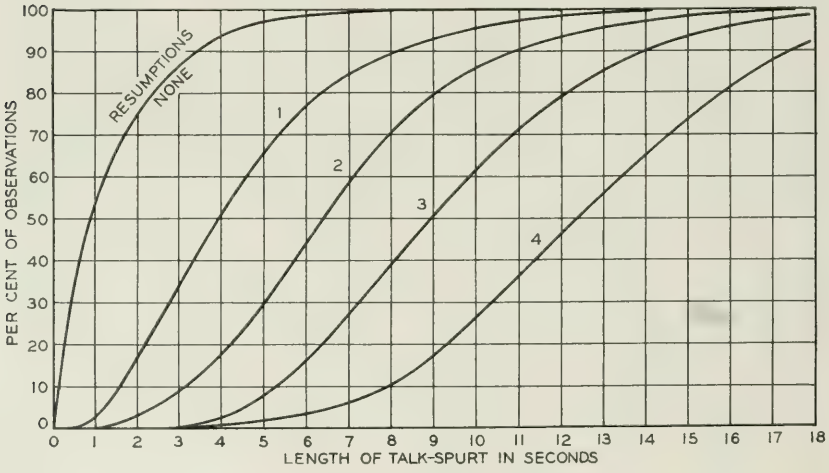


Fig. 7—Lengths of talkspurts containing a given number of pauses.

Figs. 6 and 7 show the results of analyses made to determine how frequently pauses occur within talkspurts and how the number of pauses varies with length of talkspurt. In Fig. 6 the percentage of talkspurts having a number of resumptions equal to or less than a given value is shown. It will be seen that about 60 per cent of the talkspurts contain no pauses; these comprise all the monosyllabic replies and about half the longer ones. A further analysis, shown in

Fig. 7, wherein talkspurts having a given number of resumptions are sorted according to length, indicates that almost all talkspurts exceeding 6 seconds in length contain resumptions. On the average, resumptions seem to occur about once every three and one third seconds in the longer talkspurts. The aggregate of all the resumption pauses within talkspurts amounts to about 17 per cent of the total talkspurt time.

It will be recognized that in obtaining the curves of resumption times in Fig. 4 a certain amount of arbitrary judgment must be exercised. As stated previously, the temporary absence of deflection of an oscillographic trace showing speech energy was regarded as evidence that the talker was pausing. Some assistance in determining whether or not speech energy was present was given by the trace showing the operation of the corresponding echo suppressor. The comparatively slow change in amplitude at the beginning and ending of syllables renders this determination more difficult as shorter and shorter pauses are considered. However, those shorter than about 50 milliseconds within connected speech must be observed with so much amplification that they tend to be obscured by noise on even a very quiet line. From a knowledge of the characteristics of the volume controls and echo suppressors used it is estimated that Fig. 4 represents pauses greater than 50 milliseconds during which the amplitude is lower than about 0.2 per cent of the maximum amplitude.

Conclusion

Telephonic conversation has been arbitrarily defined in terms of a few elements whose specification serves to completely describe its progress in time. These elements have been measured on a particular telephone circuit and the measurements have been presented in the form of distributions which approximate the probability of their occurrence.

A preliminary investigation under quite different conditions gave results remarkably close to those found in these more extended tests, suggesting that the conversational elements to be found throughout the aggregate of telephone users may not be materially different from those found in this investigation.

Radioactivity—Artificial and Natural [1]

By KARL K. DARROW

RADIOACTIVITY is a world-famous word relating to a world-famous subject. In general, when a statement of this sort is spoken, it is more than likely to be foolish; for specialists are altogether too addicted to imagining that their special interests are of world-wide concern when perhaps not fifty thousand people have even heard of them. With respect to radioactivity, though, the statement *is* correct. It would have been difficult indeed for any literate person to miss making the acquaintance of this word, in any year since 1900. I might say that radioactivity has been the best-advertised topic in modern physics—and this is not to say that it has been over-advertised! It is in fact the only part of modern physics which has received something approaching its due renown. Such good fortune cannot of course be altogether due to the fundamental values of the subject. A great part of the fame of radioactivity comes from medical applications and even more from medical hopes, and some from incidental things, such as the tragic end of one of its great students and the sex of two others. Still it is a matter for rejoicing that whatever the reasons may be, one portion of physics now enjoys its proper quota of the glory to which so many others are entitled.

The discovery of radioactivity took place in 1896. Quite a number of things of the highest importance to physics—and to humanity at large—were begun between 1895 and 1900, and of these the study of radioactivity was the second. The study of X-rays was the first. The second was commenced only because the first had been, and therefore I speak of the first. Imagine a tube containing air and a couple of electrodes a few inches apart, and suppose that you have a battery which can supply a durable current at ten thousand volts or more, and an air pump as well. If the air is at atmospheric pressure, the ten thousand volts can be applied between the electrodes and nothing will happen. If with the pump you now reduce the pressure of the air to about a thousandth of the atmospheric value, the air which remains in the tube will become very luminous and splendid. If next you re-

[1] A lecture sponsored by the New York Electrical Society and by the American Institute of Electrical Engineers. Delivered before the former on January 12, 1938. Scheduled for presentation to the latter at its Northeastern District Meeting in Lenox, Massachusetts, on May 19, 1938.

duce the air-pressure by yet another ten-fold, the splendor will fade out, but now the glass of the tube itself, in the region opposite the negative electrode, will be shining with a pale green glow. This too is a beautiful sight, but decidedly not one to be enjoyed for a long time at close range, for it is attended by invisible rays very dangerous to health and even life. These are the X-rays.

Many people have heard that Roentgen discovered the X-rays because he kept some photographic plates in a box which happened to lie near the tube, and found one day that the plates were fogged. It is true that he did observe the fogging of plates by the rays; it is true also that for years afterward, people were telling in various parts of Western Europe and America how they too had noticed that their plates were spoiled when left by accident in the neighborhood of discharge-tubes, but unluckily they had only felt annoyed and had resolved to keep their next batches of plates in safer places. However it was something else that led Roentgen to the discovery. He had in his laboratory some sheets of phosphorescent substances; and he found that whenever one of them was in the neighborhood of the tube it would shine; and what was more, *it would shine even when hidden from the tube by a sheet of black cardboard.* The X-rays (as he presently named them) were able to pierce substances opaque to light, and to cause phosphorescent substances to shine. Roentgen very soon discovered other properties of the rays, but these were the earliest. Moreover the greenish glowing of the glass in the X-ray tube itself resembles the light of phosphorescence; so, here were two apparent connections between penetrating X-rays on the one hand, and phosphorescence [2] on the other.

Now we come to radioactivity. There was a man in Paris whose attention was caught by these facts, and his name was not Curie. Curie is the second great name in the story of radioactivity; the first is Becquerel. I put down his first name (Henri) as well, for Becquerel—like Curie and Bernoulli and Darwin and others—is the name of a dynasty of scientists, of which Henri was the third. Henri Becquerel seems to have reasoned [3] after this fashion: "X-rays and phosphorescence are found together; therefore, wherever there is phosphorescence, there may be rays like X-rays." So he took photographic plates and wrapped them in dark paper, and then he took one phosphorescent substance after another and (after making them luminous by exposing them to light) laid them in succession beside the plates, and after an interval looked to see whether there had been fogging. Time after

[2] Or fluorescence: a careful distinction is drawn between phosphorescence and fluorescence by specialists, but need not detain us here.

[3] The idea is, however, ascribed to Henri Poincaré by Marie Curie.

time the result was completely negative, but at last he came upon a chemical compound of the element *uranium* which was phosphorescent and which fogged the plate. It is not recorded that he shouted "Eureka!" but he had as good reason so to do as Archimedes. He had discovered the first-to-be-known example of radioactivity.

Now comes the strange and paradoxical part: Becquerel had arrived at his great discovery by following a false clue. There is really no connection whatever between radioactivity and phosphorescence, and it was purely an accident that a compound which was phosphorescent had happened to contain an element which was radioactive. In trying to make a simile for what then happened, I have adopted the rather frivolous comparison which follows. Suppose that you were to meet a man who was wearing a blue serge suit, and notice that he was speaking a foreign language—Swedish, let us say. For some reason this would interest you particularly, and you would decide to look for other examples of people speaking Swedish. You would begin by reasoning that "Swedish speech was associated with a blue serge suit, and therefore any man who is wearing blue serge may speak Swedish." You would then listen to everyone whom you passed in the street who was wearing blue serge, and the first few whom you heard would prove to be speaking English. This would prevent you from believing that a blue suit necessarily entails the speaking of Swedish: but you might continue nevertheless, and eventually find another man who was wearing blue serge and speaking Swedish. Now if you were like some people, I am afraid you would send a communication to a scientific journal, announcing that it is a principle that everyone speaking Swedish is wearing a blue serge suit. But not if you were like Becquerel! If you were like Becquerel you would trail the man for weeks; and sooner or later you would come upon him wearing a grey suit or a brown one, and still he would be speaking Swedish. In the course of time you would doubtless come upon other people who never wore blue serge and yet spoke Swedish. Finally you would realize that it was just a piece of luck that you had happened to discover a man speaking Swedish, by making the fallacious assumption that all such men wear blue. Now in this case of Becquerel's, the phosphorescence was only a feature of the clothes which the uranium happened to be wearing, or more literally, the chemical compound in which it was involved. But the radioactivity was a feature of uranium itself, and that is what Becquerel proceeded to prove; first by testing various other chemical compounds of uranium which were not in the least phosphorescent, and then by testing the pure uncompounded metal itself.

Uranium, then, is a radioactive element. But it is not the only one; there were numerous others, even in the days before the physicists had started making new ones. This is where the elder Curies enter the story: Pierre and Marie Curie, in 1898. First they measured the activity of uranium, pretty carefully. Then they started measuring the activity of various minerals containing uranium, and they found too much: these minerals were more active than by virtue of their uranium content alone, they should be. The Curies suspected that some other radioactive element was lurking in the depths, and they undertook to get it out. This was of course a chemical problem primarily, and as a matter of fact, their Nobel prize was the chemistry prize, and quite rightly. Eventually they isolated their new element, or rather, two of them, which they named "polonium" and "radium." Having at last got their radium and weighed it, they found that it amounted to only two parts in a hundred million (by weight) of the rock from which they had extracted it. They required three tons of the rock, in order to get one one-hundredth of an ounce of the radium. Two parts in a hundred million! and yet, while it was still dispersed in that almost incredible scantiness through the rock, they had already been able to detect it! This is the point which I most wish to bring out, at this stage. A radioactive substance is far more easy to detect than any which is not. It is like salt in food, and the radioactivity is like the flavor of the salt, which shows the presence of a dash so insignificant that anything else in the food would go untasted if it were equally rare. Fortunately for us the instruments which are used to detect a radioactive body are not our tongues, but apparatus of a much less vulnerable kind.

Now I mention only one name before reaching the recent developments, but this the greatest name of all: RUTHERFORD. Rutherford was the first to understand radioactivity—the first to prove the contemporary atom-model—and the first to achieve transmutation. Any one of these achievements by itself would have secured undying glory to its author, but this great man made three. As lately as two years ago I wrote in a book that every leading figure in the history of transmutation was still living and still ardently at work. As lately as last October I could have repeated that, but now the Master is gone—quite suddenly gone in the fullness of his powers. This lecture is a memorial to Rutherford, not because it was expressly so designed, but because any lecture on the new radioactivity or on the old involves so much of his thought and so much of his work that it would be reduced to a few incoherent bits if everything not traceable to Rutherford should be left out.

The first achievement of Rutherford in this field was to identify the rays emitted by radioactive bodies. He identified three kinds, and gave them the names which they still bear and assuredly always will: *alpha*, *beta* and *gamma* rays. The last-named (which are of the nature of light) are the ones whereby radioactivity was first detected, for they are the ones which fog the plates and produce the phosphorescence outside of a rather narrow space just around the radioactive substance itself. They are also the ones responsible for the great work already done in medicine by radioactive bodies, and on which (I am told) there is great reliance for the future. If this were a medical lecture the gamma-rays would require most of its content; but as it is not, I leave them with this brief allusion, and turn to the others. The beta-rays are electrons, which may be of either sign (Rutherford, like the rest of the world, was acquainted only with the negative ones until five or six years ago). The alpha-rays are also charged particles, much heavier than electrons; I shall be defining them more exactly before long. The beta-rays and the alpha-rays, and for that matter the gamma-rays as well, are detected by very ingenious devices of which most types are electrical, though the particular type which supplies the photographs seen in this lecture is not.

Now at last I exhibit the list of the radioactive elements.

This list (Fig. 1) is none other than the veritable Table of the Elements itself! *Of all the known elements there now remains just one of which no radioactive form has yet been discovered or invented.* Hydrogen is the exception (the listener may say "Of course!" but it is not excluded that some day a radioactive type of hydrogen may be produced). At the end of the list stands uranium, the first of the radioactive bodies to be discovered. It has stood there ever since Mendeleieff set up the table in this fashion, but actually it now must yield its pride of place, for physicists have lately created radioactive elements which lie beyond it. I have though gone ahead too rapidly in speaking of these already. Once more let me state the marvelous fact that of all the known elements, every one but hydrogen exists in a radioactive form, or perhaps in more than one.

In speaking of "forms" I have been alluding to something which the Table as it stands in Fig. 1 does not make clear. Everyone recognizes the "chemical atomic weights" there appended to the symbol of each element. If an element has only one kind of atom, this figure is the actual mass of the atom, expressed in terms of a unit which I will presently define. There are such elements; beryllium and fluorine, sodium and aluminium are examples. Most elements, however, have two or more kinds of atoms differing in mass. Thus, in Fig. 1, the

PERIODIC TABLE OF THE ELEMENTS

(Values of atomic weights taken from the seventh Report of the Committee on Atomic Weights of the International Union of Chemistry; G. P. Baxter, *et al.*, *J. Am. Chem. Soc.*, **59**, p. 219)

I	II	III	IV	V	VI	VII	VIII			O
1 H 1.0078										2 He 4.002
3 Li 6.940	4 Be 9.02	5 B 10.82	6 C 12.01	7 N 14.008	8 O 16.000	9 F 19.00				10 Ne 20.183
11 Na 22.997	12 Mg 24.32	13 Al 26.97	14 Si 28.06	15 P 31.02	16 S 32.06	17 Cl 35.457				18 A 39.94
19 K 39.096	20 Ca 40.08	21 Sc 45.10	22 Ti 47.90	23 V 50.95	24 Cr 52.01	25 Mn 54.93	26 Fe 55.84	27 Co 58.94	28 Ni 58.69	
29 Cu 63.57	30 Zn 65.38	31 Ga 69.72	32 Ge 72.60	33 As 74.9	34 Se 78.96	35 Br 79.916				36 Kr 83.7
37 Rb 85.48	38 Sr 87.63	39 Yt 88.92	40 Zr 91.22	41 Nb 92.91	42 Mo 96.0	43	44 Ru 101.7	45 Rh 102.91	46 Pd 106.7	
47 Ag 107.880	48 Cd 112.41	49 In 114.76	50 Sn 118.70	51 Sb 121.76	52 Te 127.61	53 I 126.92				54 Xe 131.3
55 Cs 132.91	56 Ba 137.36	RARE EARTHS	72 Hf 178.6	73 Ta 180.88	74 W 184.0	75 Re 186.31	76 Os 191.5	77 Ir 193.1	78 Pt 195.23	
79 Au 197.2	80 Hg 200.61	81 Tl 204.39	82 Pb 207.21	83 Bi 209.00	84 Po	85—				86 Rn 222
87—	88 Ra 226.05	89 Ac	90 Th 232.12	91 Pa 231	92 U 238.07					

RARE EARTHS

57 La 138.92	58 Ce 140.13	59 Pr 140.92	60 Nd 144.27	61	62 Sa 150.43	63 Eu 152.0	64 Gd 156.9
65 Tb 159.2	66 Dy 162.46	67 Ho 163.5	68 Er 167.64	69 Tm 169.4	70 Yb 173.04	71 Lu 175.0	

Fig. 1

pigeonhole for hydrogen should contain three mass-values; that for helium, two; that for tin, no fewer than ten. It would make the table impossibly crowded to print them all in this way, and consequently I have broken it up into sections, of which Fig. 2 represents the first six elements.

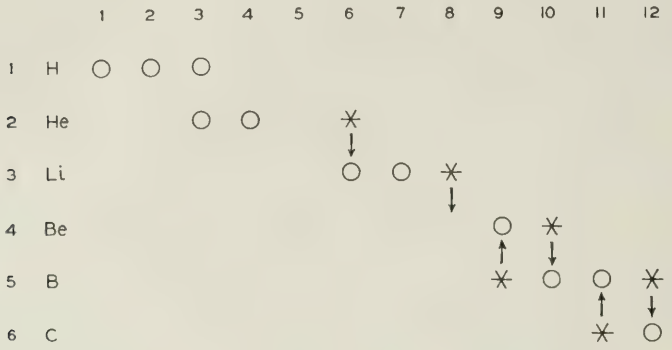


Fig. 2—Isotopes of the first six elements.

In this figure each element has a row to itself, and each value of mass has a column to itself, and each circle represents a stable kind of atom. I now introduce the technical term "isotope" to distinguish the different kinds of atoms common to a single element. Hydrogen, you see, has three stable isotopes (there is some doubt about the stability of the third, though none about its existence); helium two (again there is doubt about the stability of one); lithium two, beryllium only one, boron two, and carbon two of which the second will appear in the next figure. The unit of mass is a very small amount, about $1.67 \cdot 10^{-24}$ of one gram. I do not pause to give it as accurately as I might, for we are not going to be concerned with very exact mass-values in this talk. The masses of the isotopes are not exactly integer multiples of this unit; for instance, those of the three kinds of hydrogen atoms are 1.008, 2.016 and 3.017. The departures from integer multiples are, however, small, as you see in these three cases. Small as they are, they are mightily important; but it is permissible to ignore them for the purposes of this lecture, and I am going to ignore them from now on. I will, however, speak of the integers at the heads of the columns as "mass-numbers" rather than "masses."

Since the isotopes of an element differ in mass, what is it then that they have in common? I answer this question by describing Rutherford's second great achievement, the "nuclear atom-model." Rutherford was the first to prove that the atom consists of a positively-charged nucleus surrounded by a swarm of negative electrons. The

nucleus is much more massive than the electrons, and this is one of the reasons for comparing the atom with the solar system, in which the sun is much more massive than the planets which perpetually swing in orbits around it. A less hackneyed and newer simile is that of Bragg, just now, by the way, appointed as Rutherford's successor in the Cavendish chair at Cambridge, who likens the atom to a man's head with a swarm of gnats buzzing around it. Normally—that is to say, when the atom is complete and electrically neutral—the negative charges of all the electrons put together just balance the positive charge of the nucleus. If Z be used to stand for the number of electrons in the normal neutral atom, and $-e$ for the charge of the negative electron, then $+Ze$ is the amount of the charge on the nucleus. Z is called the "atomic number" of the element in question.

Fig. 3—Isotopes of the elements numbered 6 to 11.

This it is, of which all the isotopes of any one element have the same value. This it is which distinguishes an element, and is common to all of the different atoms of that element whatever their masses may be. For hydrogen it is 1; for helium 2; for lithium 3; for uranium 92. Each row in Fig. 2 and Fig. 3 is marked on the left not only with the chemical symbol of the element to which the row belongs, but also with the atomic number thereof.

Radioactivity is a feature of the nucleus. This accounts for some of its remarkable aspects, which greatly surprised the world of physicists and chemists when they were first established. Being a quality of the nucleus, it varies from one isotope [4] to another of any element, much more drastically than does the mass. Being a quality of the nucleus, it is immune to the physical state of the atom—*i.e.* it is the

[4] Isotopes were first distinguished from each other (by Soddy) by virtue of their differences in radioactivity.

same whether the atom is part of a solid, a liquid or a gas; it is immune also to the chemical state of the atom, *i.e.* it is the same whether the radioactive element is isolated or is part of a chemical compound. It is also immune to heat and to all of the many other agencies which physicists and chemists have at their command.

Radioactivity being a feature of the nucleus, every chemical symbol which I use from now on will refer to the nucleus of an atom and not to the atom as a whole. "Be" will stand for beryllium nuclei, "F" for fluorine nuclei, "Al" for aluminium nuclei. For most elements, though, there are two or more different sorts of nuclei distinguished from each other by their masses, and the symbol must tell us which is meant. The custom is to write the mass-number of the isotope in question as if it were an exponent: H^1 and H^2 and H^3 for the three kinds of hydrogen nuclei, He^3 and He^4 for the two kinds of helium nuclei, Li^6 and Li^7 to distinguish between the isotopes of lithium, and so on. If in addition one wants to remind the reader of the atomic number, one writes it as a subscript before the chemical symbol: ${}_1H^1$, ${}_1H^2$, ${}_2He^4$, ${}_9F^{19}$ and the like. Purists object that either the chemical symbol or the value of Z is superfluous when both are given, but others often like to see them both. And now for some names: there are three nuclei which have names of their own. The Greek words for "first" and "second" are applied to ${}_1H^1$ and ${}_1H^2$; they are the *proton* and the *deuteron*. The name for ${}_2He^4$ is *alpha-particle;* this nucleus is indeed the particle which, as Rutherford discovered long ago, makes up one of the three kinds of rays which radioactive bodies emits, and there never was a greater piece of good fortune in language than that whereby this all-important particle received the name of the first letter of the Greek alphabet, for indeed it is the alpha of modern nuclear physics. And now another reference to masses: the mass of the electron (when not moving extremely fast) is only about .0005 of the mass-unit which is being used throughout this talk, and therefore the mass-numbers at the heads of the columns in Figs. 1 and 2 and others are about as good approximations to nuclear masses as they are to atomic masses, and I shall use them as such.

Now let us notice not only the circles of Figs. 2 and 3, but the stars as well. The stars also stand for nuclei, but these are *radioactive*— or *unstable*, two words which have practically the same meaning when applied to a nucleus. At least one star appears in every row, the first in Fig. 2 excepted. If the figure had room for ninety-two rows, one for each element from hydrogen to uranium, there would appear at least one star in every row below the first, excepting three (atomic numbers 61, 85, 87) for which no isotope either stable or unstable

known with certainty and the element itself must still be regarded as missing. (Furthermore there should be at least three more rows numbered 93, 94 and 95, and containing stars but no circles.) This is what is meant by saying that every known element, hydrogen alone excepted, has at least one radioactive form.

Figure 2 shows that at the beginning of the Table of the Elements, the stable types of nuclei outnumber the unstable ones. The preponderance is gradually shifted as Z increases, and Fig. 4 exhibits to us how greatly the radioactive nuclei outnumber the stable ones among the elements of which the atomic numbers range from 81 to 84. Indeed the circle which is lowest and most to the right in Fig. 4 represents the most massive and most highly charged of all the stable nuclei which are known (it is the solitary isotope of bismuth, atomic number 83 and mass-number 209). All the rows after 83 are occupied entirely by stars.[5]

Fig. 4—Isotopes of the elements numbered 81 to 84.

As the title of this talk has already suggested, the radioactive nuclei are of two classes: the "natural" and the "artificial," the types already existing in the rocks of the earth and the types made in the laboratory by physicists employing the art of transmutation. Nearly all of the natural types lie beyond 80 in atomic number, and most of them were discovered in the first fifteen years after Becquerel found the first. Two of them are identical with two of the man-made types. Apart from these two, every one of the artificial types is a creation of the years since 1933. One guesses that while the natural radioactive bodies may be many, the artificial ones must surely as yet be few; how surprising then to learn that while there are some forty of the former, the latter after four brief years already number two hundred and thirty! Unlike the natural ones, these artificial isotopes are sprinkled liberally throughout the whole of the Table of the Elements, from the second onward to the end. Not only in number but in di-

[5] It must though be admitted that some of the heaviest nuclei, though demonstrably radioactive, may exist for hundreds of millions of years before they disintegrate; "instability" is indeed a very relative concept!

versity of mass and charge are the artificial radioactive bodies now outstanding by far.

These artificial examples forming now so large and important a group among the radioactive nuclei, I make a digression to speak of some of the transmutations from which they are derived. The art of transmutation is already so huge a subject that the digression must be severely limited if ever we are to come back to radioactivity. I must therefore make only a passing allusion to the fact that the first of the new radioactive nuclei were made by bombarding various light elements with very energetic alpha-particles. Here the second Curie generation must be introduced, for the daughter and son-in-law—Irene Curie and Frederic Joliot—of the first Curie pair were the ones who made this discovery. (It was not their entry into the field of radioactivity, they having already studied natural radioactive bodies for a number of years).

Returning to Figs. 2 and 3, notice that many of the radioactive isotopes lie just one step to the right of stable isotopes: Li^{8}, Be^{10}, B^{12}, C^{14}, N^{16}, O^{19}, F^{20}, Ne^{23} are the examples found in these two pictures alone. It seems as though they might differ from their neighbors on the left—Li^{7}, Be^{9} and so forth—by possessing an extra particle of mass (approximately) 1 and charge zero. If only one could find such particles roaming freely about in Nature, might one perhaps succeed in adding them to the stable nuclei of lithium and beryllium and boron and the other elements, and so produce these radioactive nuclei?

Such particles may indeed be found roaming about in Nature, but not of their own volition. These "neutrons"—for such is their name—must themselves be set free by the art of transmutation. Free neutrons were first produced by bombarding certain elements with alpha-particles; the discovery was an international one, and its story is interesting, but to keep this digression within bounds I must again content myself with giving the names—Bothe and Becker in Germany, Curie and Joliot in France, Chadwick in England—of those who carried it through its consecutive stages from first intimation to triumph. More than a hundred different ways of freeing the neutron are already known, but of all this diversity I will take one only, which consists in projecting deuterons against deuterons.

The "deuteron-deuteron reactions"—D-D reactions for short—are produced by applying high voltage to deuterons (emerging from a discharge-tube containing heavy hydrogen, in which some of the atoms are divested of their electrons and the nuclei are left bare) and then directing them across a vacuum against a target containing other deuterons. (The target may be some solid compound of heavy

hydrogen, such as ice in which plenty of the hydrogen atoms belong to the isotope H^2; or it may be gaseous heavy hydrogen). The high voltage is required, so that the impinging deuterons may override the electrostatic repulsion between the positive charges which they bear and the positive charges of the deuterons waiting in the target, and come into contact with these last. Generally in transmutation, "high voltage" signifies volts by the millions. These particular reactions are, however, among the easiest to produce, and with less than a hundred thousand volts it is quite possible to liberate neutrons at such a rate that their peculiar qualities can be well studied. (One reaction indeed has been detectably produced at 8000 volts, a figure so low that it arouses speculation as to what the course of physics might have been if the second isotope of hydrogen had been discovered say thirty years ago.)

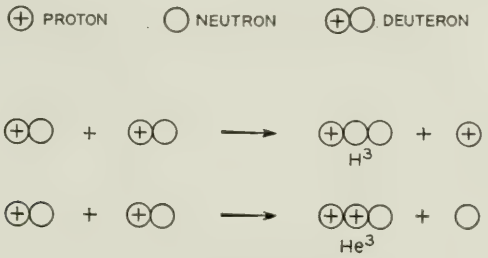


Fig. 5—Scheme of the deuteron-deuteron reactions.

To explain what is actually observed to happen, I ask the listener to imagine the deuteron as a composite of a proton and a neutron, as it is exhibited in Fig. 5. With this image in mind, one might well expect that when deuterons are hurled with great energy and speed against a plate of matter containing massive nuclei—lead, for instance—they would be broken in two. This has been sought for but apparently does *not* happen, showing that we must keep our imaginations under continual check by experiment. What *does* happen is displayed, for the impacts of deuteron against deuteron, in Fig. 5. It seems that one deuteron is after all broken in two, but only under the condition that either its component proton or its component neutron adheres to the other. Another metaphor: one deuteron snatches either the proton or the neutron away from the other, leaving the abandoned neutron or proton to go free. Both of these descriptions are too figurative, but what is certain is this: from the scene of such impacts, particles of all the four kinds shown to the right of the arrows in Fig. 5 are observed to be proceeding. The labels show (what should already be obvious) that the newborn particles of mass 3 are isotopes

of hydrogen or helium, according as they contain two neutrons and one proton or two protons and one neutron. I now show pictures to support these statements.

In Fig. 6 the apparatus is shown in a sketch: the cloud-chamber or expansion-chamber of C. T. R. Wilson, being the hollow cylinder which is shown below in axial section, its top being a glass plate and its bottom a piston-head which can be pulled very suddenly downward

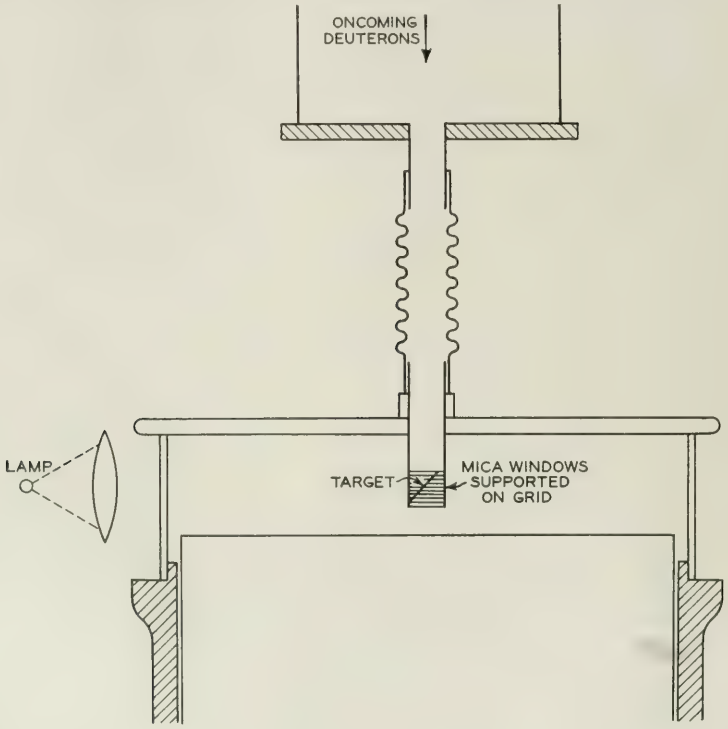


Fig. 6—Expansion-chamber arranged for detecting transmutation by deuterons.

by mechanism. Ordinarily the chamber is filled with moist but dustless air; when the piston-head suddenly drops, the air and the water-vapor are sharply cooled by expansion, and the vapor condenses in droplets upon whatever ions may be floating in it. The side-tube which enters the chamber from above is evacuated; through it come the impinging deuterons, to make their impacts upon the target at the knob-like closed end of the tube. The wall of the tube, thin as it may be made, is too thick to allow the deuterons to emerge into the air of the chamber. One might well expect that *a fortiori*, any new particles born out of the transmutation would be too slow-moving to

pierce the wall; but many of these particles are much more energetic than the impinging deuterons themselves, for they draw upon a reserve of energy stored up in the nuclei.[6] They shoot through the wall into the air of the cloud-chamber itself, and if they are charged, they make long trails of ions along their paths. The expansion is then produced and the water-vapor, condensing upon these ions, makes trails of droplets which are the paths made visible.

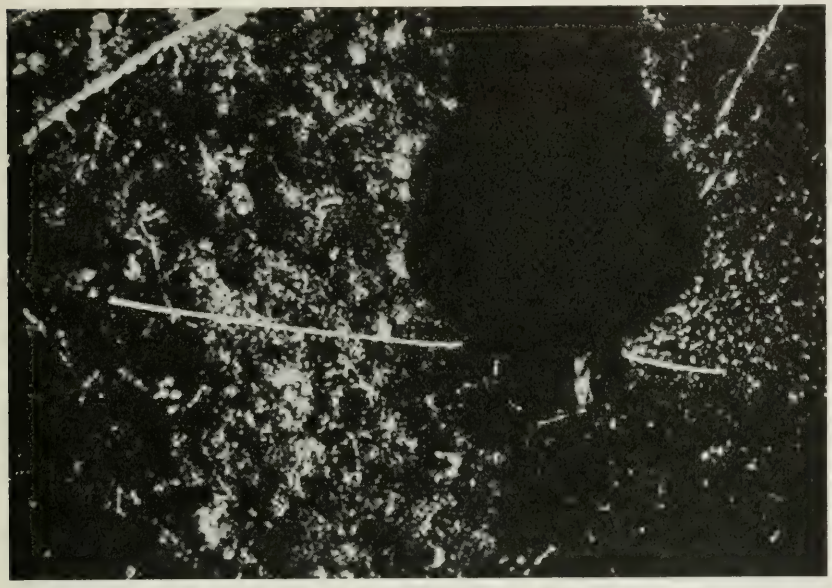


Fig. 7—Tracks of a proton and a H^3 nucleus resulting from one of the deuteron-deuteron reactions. (P. I. Dee, Cavendish Laboratory; *Proceedings of the Royal Society.*)

Figure 7 exhibits two of these visible paths of "tracks," made by particles which sprang from the scene of the transmutation (inside the knob) in practically opposite directions but with very different penetrative powers, since one of the tracks is seen to be much longer than the other. The long one is the track of a proton, the short one is that of a H^3 nucleus which is a deuteron augmented by a captured neuteron; this picture shows a single example of the upper reaction of Fig. 5. How can physicists be sure that these tracks are due to the nuclei which I have named? This question is far too deep to be answered in this place, and I can only assure the listener that while such pictures by themselves cannot suffice for the proof, an unassailable

[6] In the language employed in chemistry, these are "exothermic" reactions.

proof can be and has been given by other and electrical methods of observing these newborn particles.

But how about the lower reaction of Fig. 6—the one which really concerns us, since all this digression is designed chiefly to exhibit the origin of free neutrons? In Fig. 7 no track appears which can be attributed to either a He^3 nucleus or a neutron; and no such tracks appear in other similar pictures. The absence of He^3 is, however, due to a simple cause; these nuclei are born with insufficient energy to traverse the wall of the tube. To observe their tracks it is necessary to suppress the tube-end, to fill the expansion-chamber with heavy

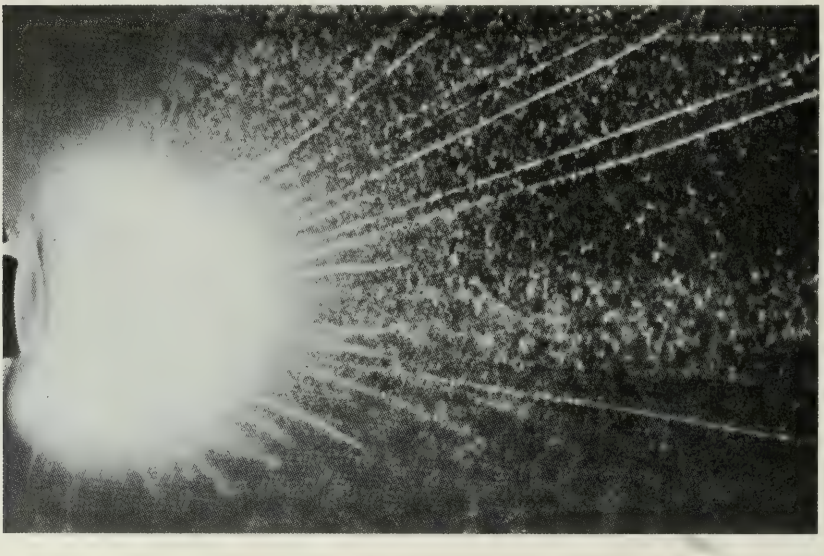


Fig. 8—Tracks of protons, H^3 nuclei and He^3 nuclei resulting from the two deuteron-deuteron reactions. (P. I. Dee and C. W. Gilbert, Cavendish Laboratory; *Proceedings of the Royal Society.*)

hydrogen in the gaseous form, and to project the deuterons directly into it. When this is done, all of the region of the gas which the impinging deuterons can reach becomes completely filled with the ions formed along their many tracks, and appears as a flare on the photograph (Fig. 8). Out of the flare project the tracks of the newborn nuclei. Those which stretch clear across the picture are in part those of protons, in part those of H^3 nuclei born from the first reaction. But in addition one sees a number of short tracks which terminate not far from the edge of the flare itself. These are the tracks of He^3 nuclei—not merely guessed, but proved, to be such.

Where, however, are the tracks of the neutrons? They are not seen upon this picture, nor in any; for neutrons make no tracks. Neutrons bear no electric charge, and hence they do not ionize the molecules of the air or any gas as they go through, for ionization is effected only by electrical forces which can tear electrons out of their places in molecules. Not making ions, they afford no footholds whereby the water-vapor can condense and mark their passage. The expansion-chamber is frustrated; and worse yet, so are the electrical devices which serve for detecting charged particles like fast protons or fast electrons, since they, too, depend on the ions which these can make. The neutron indeed might slip through all of our apparatus completely undetected, were it not liable to make collisions with nuclei so sharp and sudden

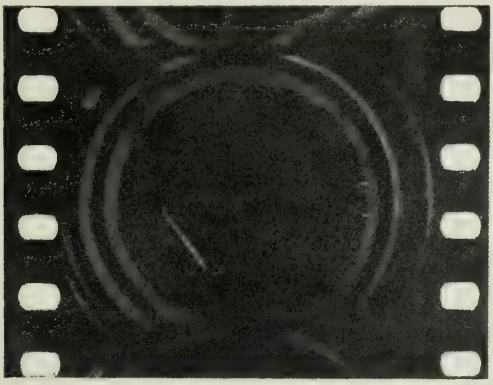


Fig. 9—Track of a proton recoiling from the impact of a neutron. (I. Curie and F. Joliot, Institut du Radium; *Journal de Physique.*)

that they might be compared with impacts of one billiard-ball upon another. Comparisons with billiard-balls are rife in physics, but seldom with so much justification. When neutrons are streaming through a gas, such impacts are suffered by nuclei of occasional atoms of the gas; and like a struck billiard-ball they recoil, and in recoiling they are able to make ions and the ions then serve to reveal them.

The track of such a recoiling nucleus, made visible in a cloud-chamber, is seen in Fig. 9. This was taken soon after the discovery of the neutron, and at a time when these particles had as yet been released only by using natural radioactive substances to project alpha-particles against various targets. Such ways of producing free neutrons are not very efficient, and accordingly one sees in the whole expansion-chamber one track and one only (though I must interpolate that according to our present knowledge, many thousands of neutrons

must have traversed this chamber without happening to strike any nucleus). See now the contrast with the present time, as illustrated by Fig. 10 which shows an expansion-chamber traversed by the

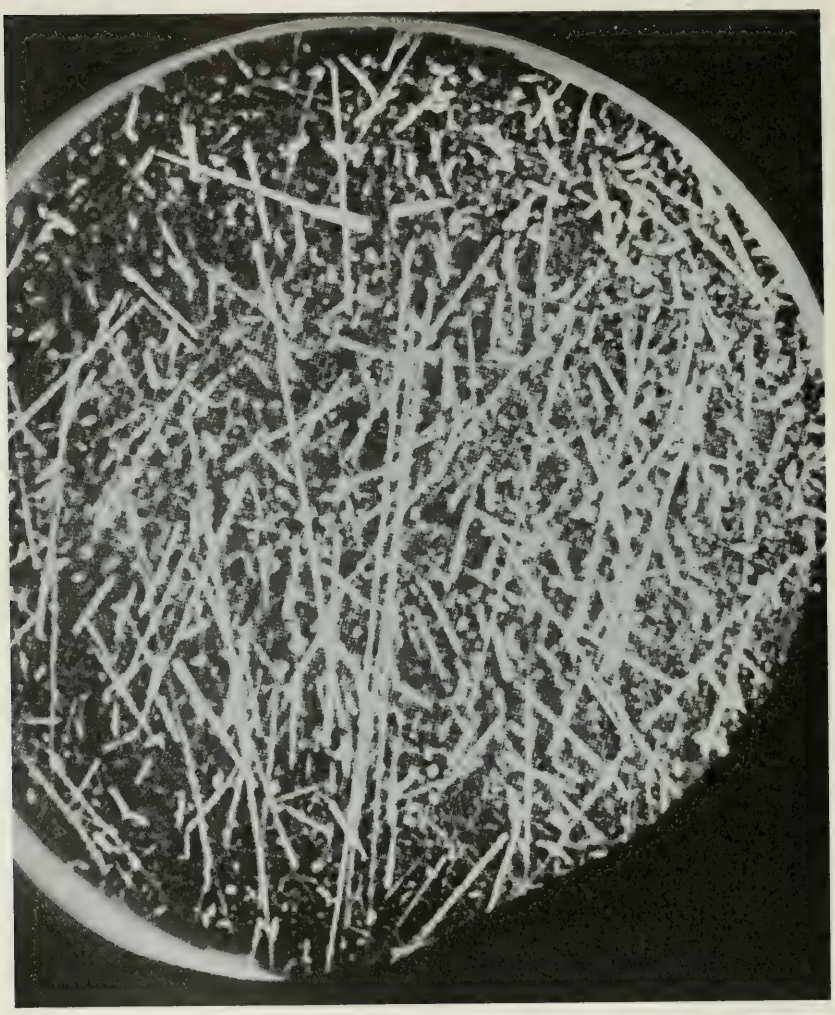


Fig. 10—Tracks of protons recoiling from impacts of a dense stream of neutrons. (F. N. D. Kurie, University of California.)

neutrons released when deuterons from a high-voltage machine bombard a target.[7] The machine in question was the famous cyclotron of E. O. Lawrence; there is no more striking illustration of the powers

[7] This was another reaction than the one just mentioned, the target being of beryllium.

which this instrument has conferred upon the scientific world, over and above those which radium has already granted.

Our digression now ends, and we return to the artificial radioactive substances, being now equipped with knowledge as to how these—or rather, many of them—can be made. As I said earlier, many radioactive isotopes differ from existing stable isotopes only in possessing an extra neutron in the nucleus; and this extra neutron can be supplied to the stable nucleus, combining it and converting it into the radioactive type. I have just exhibited one way in which the extra neutron may be, and often is, supplied. In the first-mentioned of the deuteron-deuteron reactions, a neutron is taken away from one of the deuterons by the other, which latter is thus converted from H^2 into H^3. There are many stable isotopes, of many elements, which are able to take away neutrons from impinging deuterons in this manner; a recent list gives no fewer than fifty. The resulting nucleus-types are not in every case radioactive; several are stable, including H^3 itself (at least, no one has yet discovered evidence that H^3 is unstable, though there are doubts about it). Most however *are* radioactive. The reactions in question are known as (d, p) reactions, in allusion to the fact that deuterons enter the target and protons spring out. One might imagine that the deuteron consists of a proton leading along a neutron, which it pushes into the nucleus which it strikes, itself continuing its career as a free particle.

Neutrons, however, do not have to be escorted into nuclei by protons; those which are already free, such as the ones which are released in the second of the D–D reactions, are quite well able to creep in themselves and make themselves permanently at home. My use of the verb "creep" is not entirely fanciful, for the slower the neutrons are moving as they approach a target, the better their chance of entering its nuclei. Those fresh from their origin in reactions of transmutation are usually moving much too rapidly to be able to come to a halt in a nucleus—or to be liable to capture, whichever way of putting it one may prefer. It is necessary to interpose, between the source and the target, a block of paraffin or a can of water several inches thick. If the source consists of a natural radioactive substance bombarding another element with alpha-particles and thus releasing free neutrons, the two may be mixed with each other and enclosed in a capsule which is then embedded in the centre of a paraffin sphere or immersed in water. As the neutrons make their way out, they collide again and again with the nuclei of atoms in the paraffin or the water, and these recoil from the impacts. It is not, however, their recoiling which is now of importance, but the fact that at every

such impact the neutron loses some of its energy. The point about choosing water or paraffin is that they are rich in hydrogen and consequently full of protons, and the elastic impacts of the neutrons against these entail a greater average loss of neutron-energy per collision than do impacts against any other nuclei.[8] There is good reason to believe that most of the emerging neutrons have energies no greater than those which the atoms of the water or the paraffin possess by virtue of their thermal agitation. These are the neutrons which are most effective in converting stable into radioactive nuclei by letting themselves be captured.

Even yet I have not mentioned all of the ways in which radioactive substances can be and are being made. Time does not suffice for commenting on the others, but some are exhibited in Fig. 11, which

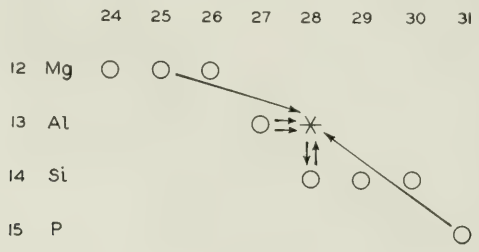


Fig. 11—Various ways of making the radioactive nucleus Al^{28}.

displays all of the stable but only one of the unstable isotopes of the four elements numbered 12 to 15. This one radioactive type, which I have tried to make more conspicuous by leaving out the rest, is the isotope 28 of aluminium—one of the few elements, which, very conveniently for physicists, has one stable isotope only. The arrows converging onto the star show the different ways in which Al^{28} is made. The two coming from the left signify that it is made by adding a neutron to the stable isotope Al^{27}; there are two of them, because the Al^{27} nucleus can either absorb a slow free neutron or annex the neutron from an impinging deuteron, whichever it has the opportunity of doing. The arrow slanting downward from the left signifies that Al^{28} can be made by bombarding magnesium with alpha-particles; some of these are absorbed by nuclei of the isotope Mg^{25} which thereupon at once emit protons. The arrow slanting upward from the right signifies a process in which phosphorus is bombarded by neutrons (fast ones, in this case!) and some are absorbed by the nuclei P^{31}

[8] Energy-transfer between an initially-moving and an initially-stationary elastic sphere is greatest when the latter is of the same mass as the former, and for protons and neutrons this equality of mass is realized within 0.1 per cent.

which thereupon eject an alpha-particle apiece. The arrow pointing straight up signifies a process in which silicon is bombarded by neutrons; some are absorbed by nuclei Si^{28}, which instantly throw out protons. With no fewer than five ways of making a single radioactive type at his command, the physicist is in a position of power which seems all the more remarkable when one recalls that as lately as five years ago he had not (knowingly) made any radioactive substance by any way whatever.

Consider now the arrow pointing away from the solitary star in Fig. 11, and the arrows pointing away from the many stars in Figs. 2 and 3. These signify what is really meant by calling an isotope "radioactive." *A radioactive nucleus is one which spontaneously changes itself into a nucleus of another element by emitting a charged particle.* (Usually it lasts an appreciable time before it does so, and this delay is to be mentioned in a complete definition of the word "radioactive.") The arrows pointing away from the stars will serve to specify these changes. All in these figures are vertical; every one of these unstable isotopes transforms itself by emitting a particle of which the mass is very small (compared to the mass-unit which we are using) while the charge is $+e$ or $-e$, according as the transformation is to the element preceding or to the element following. These particles are positive and negative electrons. All of the man-made unstable nuclei are radioactive after this fashion, being electron-emitters; and so are more than half of those which are found in Nature.

What decides whether it shall be a positive or a negative electron which a given nucleus-type emits? Physicists cannot explain this as yet, in any adequate sense of the verb "to explain"; but we can readily see the law which governs the choice by examining the pictures. In Figs. 2 and 3 it will be seen that from each star the arrow points in whichever sense—upward or downward—it finds a circle to point at. Becoming completely animistic for the moment, I may say that the unstable nucleus wants to be stable—knows that one of its two neighbors, of identical mass-number but greater or lesser charge, is stable—knows *which* of the two is stable—and deliberately proceeds to identify itself with its stable neighbor by emitting an electron of the necessary sign. Putting the situation more drily: each of these unstable nucleus-types tends to transform itself into its adjacent stable isobar. Here "isobar" is a technical term for "nucleus of the same mass-number," and "adjacent" is a short way of saying "belonging to the preceding or the following element."

Suppose the star has circles both above and below it, *i.e.* that both of the adjacent isobars are stable (and prove themselves so by existing

in Nature): how will the unstable nucleus resolve its dilemma? Such cases are rare, but not entirely absent. An example appears in Fig. 12. The elements palladium and cadmium have isobaric isotopes of mass-number 106, in spite of the fact that their atomic numbers (46 and 48) are not consecutive; silver, with atomic number 47, lies between. There is no stable silver isotope 106, but a radioactive one can be and has been created, and for this the dilemma is posed. It handles the

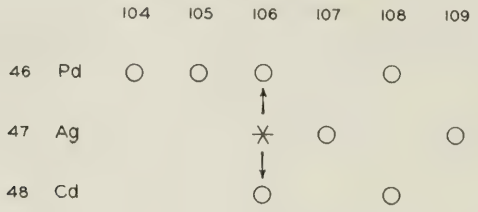


Fig. 12—Illustrating an example of isomers.

dilemma by grasping both horns! electrons of *both* signs come out of the radioactive silver. I must say that there is something which indicates that the nuclei which make one choice may be slightly different (in mass, for instance) from those which make the other. It may therefore be well to speak of silver as having two isotopes of the same mass-number 106, and a word has already been coined: they are called "isomers" of one another. This does not alter the fact that where alternative choices exist, both are elected.

On Fig. 2 we notice an arrow which points to a vacancy. No stable nucleus Be^8 is known, though there has been a very diligent search for it conducted by many ways in many places. Practically no doubt exists that Be^8 bursts of itself into two pieces (two alpha-particles) almost as soon as it is made. We thus have here an unstable (radioactive) nucleus—Li^8—which does *not* find stability by ejecting an electron, but instead hastens onward to a completer ruin. In the lower reaches of the Table of the Elements there are so many stable isotopes that the unstable ones can almost always turn themselves by one electron-emission into one or another of these, and such catastrophes are rare. Among the natural radioactive substances in the upper reaches of the Table they are common, as I now show in returning to natural radioactivity for the close of this talk.

Notice again Fig. 4, in which the stars are so many and the circles so few. If arrows were to be inserted to show the transformations, they would crisscross into a maze. I have therefore separated the figure into three: all of the circles, stars and rosettes in it will be found

in Fig. 13 (which is really a pair of figures, as the caption says) and Fig. 14.

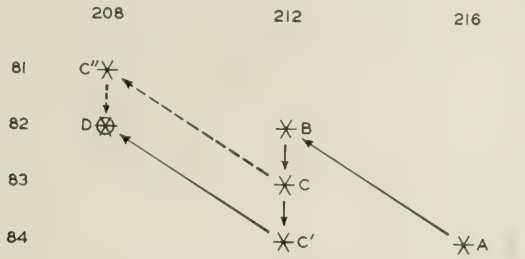


Fig. 13—Part of the thorium series of radioactive nuclei. To obtain the actinium series, imagine each star and rosette transposed one unit to the left.

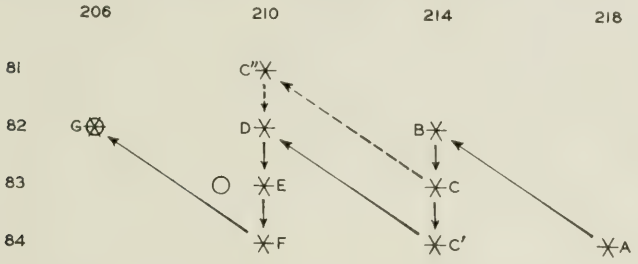


Fig. 14—Part of the radium series of radioactive nuclei.

Taking Fig. 13 as it stands, we see five of the stars and one rosette of Fig. 4 connected by arrows; each star marks a nucleus which transforms itself by emission of a particle into the one next following along the arrow-chain. The rosette stands for a stable nucleus, and would be replaced by a circle were it not distinguished by being the terminus of such a chain. These five radioactive isotopes and one terminal nucleus-type belong to the "thorium series," and are known as thorium A, thorium B, thorium C′ and so on, according to the letters which adjoin their stars. This is an unlucky bit of terminology, for it suggests that all are isotopes of the same element, which is clearly far from the truth.

Taking Fig. 13 again and imagining each star moved by one unit to the left (so that *e.g.* the star A goes to 217), we now are thinking of five more stars and another rosette of Fig. 4, duly connected by arrows. These constitute (a part of) the "actinium series" and are known as actinium A, actinium B and so forth.

Taking Fig. 14 as it stands, we find ourselves confronted by eight more of the stars and the last rosette of Fig. 4, connected by arrows. These constitute a part of the "radium series" and are known as

radium A, radium B and so on. The "so on" covers more than it did in the other two cases, this chain continuing to the terminus here marked as radium G, though usually known by a different name.

Surveying the scene of these *massive* radioactive nuclei, one is struck by the fact that not all of the arrows are vertical. Many are slanting, and by their slant and their length they show that they represent the emission of alpha-particles. It is a feature of some (not of all) of the unstable nuclei of mass-numbers greater than 200, that they strive toward stability by emitting these. For this feature we should be very grateful, since it was by the use of alpha-particles from natural radioactive bodies that Rutherford achieved the first of transmutations; though physicists now can transmute without their aid, no one can guess how long they would have waited without trying had they not had that encouragement. Vertical arrows also are seen, but again there is a contrast to the lighter isotopes; all of the electrons emitted by radioactive nuclei of mass-numbers beyond 200, or by natural radioactive isotopes of whatever mass, are negative. But for the fact that positive electrons had been observed among the cosmic rays in 1932, they would have been discovered along with the first examples of artificial radioactive isotopes in 1934, and what a sensation that would have been!

More than by anything else, probably, one is impressed by the concatenation of these radioactive nuclei. A long journey to stability lies ahead of thorium A and actinium A, a longer one still ahead of radium A; but the total lengths of the journeys are greater yet, for they begin farther back. In Fig. 15 we behold the three series of radioactive isotopes in their entirety, and it is seen that the three "A-products," as they are called, are midway in the evolution and not at its beginning. The manner of drawing of this figure is changed from the preceding, atomic number being laid off along the horizontal axis and mass-number along the vertical; also, crosses and circles and dots are used to mark the members of the actinium, radium and thorium series respectively, and have no bearing on stability or instability. The actinium series should lie lower than it is drawn, with its terminus AcD lying midway between ThD and RaG; the mistake is incurred so as to diminish the overlappings which would otherwise confuse the picture.

Except for a few created in the last three years by transmutation, every known nucleus-type of mass-number greater than 209 and atomic number greater than 83 (as well as a few of slightly lower values) is found in Fig. 15. It appears that 83 and 209 are critical values of nuclear charge and mass, beyond which the constituents of

nuclei—neutrons and protons, presumably, and whatever others there may be—cannot unite into stable systems.[9] All of these nuclei beyond 209 which are found in Nature are seeking for stability by the emission of particles, but never finding it until they have emitted

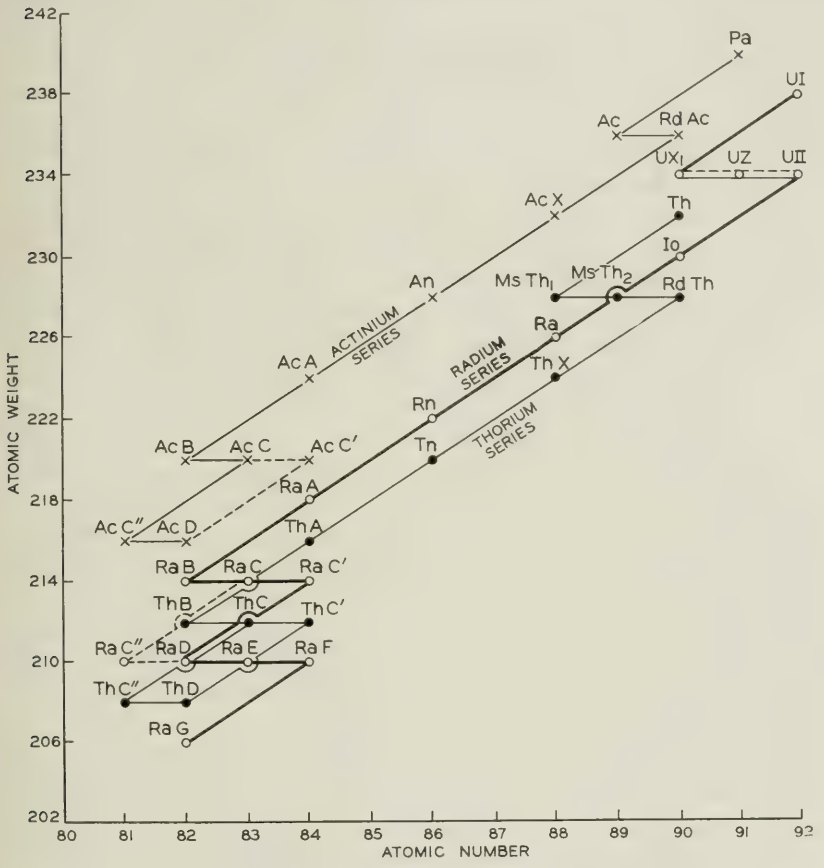


Fig. 15—The three series of radioactive substances.

sufficiently many to convert themselves (or rather, the residues of themselves) into one or another of the three isotopes of element 82—*lead*—which are marked by rosettes in Fig. 4. For an obvious reason these three are called thorium lead, actinium lead and radium lead. "Ordinary" lead as it comes from most mines is a mixture of many isotopes, but the lead which is found in close association with thorium or with uranium proves its origin from vanished atoms of these metals

[9] I recall from an earlier footnote that some of these very heavy nuclei (notably thorium 232 and uranium 238) are so very long-lasting that "unstable," while strictly correct, seems much too strong a word for them.

by being preponderantly the isotope 208 or the isotope 206 as the case may be.

I pause to mention, in justice to Rutherford, that it was he who proved by study of some of these natural radioactive bodies that each is transforming itself into a different element; also it was his associate Soddy who by similar studies was led to distinguish the first-to-be-recognized isotopes, that is, different radioactive forms of one and the same element. The way for making such diagrams as Figs. 2 and 3 and 4 was prepared before 1914 by these two men, though some of the facts embodied in Fig. 4 were not then available because of want of knowledge of atomic numbers, and all of the knowledge embodied in Figs. 2 and 3 was non-existent because no radioactive isotopes of these elements had yet been created and nobody knew as yet how to distinguish their stable isotopes. Also it should be mentioned that only the extraordinary potency of radioactive substances in affecting our instruments of measure enables the physicist or the chemist to recognize the element to which a radioactive isotope belongs, nay even to detect its presence. With radium and a few others, it has been possible to amass enough of the substance to see and to weigh; with the great majority of natural and with the totality of artificial radioactive isotopes, nothing of the sort has even been approached, and we should still be unacquainted [10] with them if they had been stable.

Three examples of transmutation which occur in these upper ranges of the Periodic Table deserve to be recorded in even so brief a report.

In Fig. 14, notice the circle in row 83 and column 209 which (as I earlier said) represents the highest stable nucleus—bismuth 209. Radium E, represented by the star to its right, is clearly bismuth 210; by the testimony of its mass-number and atomic number, it differs from stable bismuth nuclei by the possession of an extra neutron. If bismuth should be bombarded by neutrons, either free or bound into deuterons, would it be transformed into radium E? Livingood at Berkeley did bombard ordinary bismuth with very energetic deuterons, and did succeed in producing a radioactive substance which agreed with radium E not only in emitting negative electrons, but also in converting itself into a substance emitting alpha-particles, and the agreement extended to details of the emission. No doubt exists that he was making radium E out of bismuth 209 by enabling deuterons to transfer their constituent neutrons to this latter, just as H^3 is made from H^2 in the first of the deuteron-deuteron reactions.

[10] This statement should be qualified slightly, for some of the artificial radioactive nuclei spring from reactions of transmutation which are so well understood that the observer could justifiably infer the existence of the nuclei in question even if he did not observe them.

In Fig. 15, notice that all of the members of the thorium series have mass-numbers divisible by 4, or equal to $4n$ with various integer values given to n. This is accordingly called the "$4n$ series," and one readily sees what is meant by calling the radium and the actinium series by the names of "$4n + 2$" and "$4n + 3$" series respectively. One begins at once to wonder whether there is not a "$4n + 1$" series. Such a series was long sought after in vain, and no member of it has yet been discovered in Nature; but in the laboratory of the Curies in Paris thorium has lately been strongly bombarded by neutrons, and a new sequence of radioactive bodies has thus been engendered which has already been followed through several steps, and is in all probability the series so long missing.

As to the remaining feat—the creation of elements beyond uranium —it is now beyond doubt. Fermi and his school at Rome, Hahn and Meitner in Berlin, Curie and Joliot in Paris have all borne witness to it. In one way it seems the most romantic of all the feats of transmutation, for Nature had apparently set 92 as the limit of nuclear charge, and now man has transgressed it. The process is begun by exposing uranium to bombardment by streams of neutrons. It appears that when a uranium nucleus has captured a neutron, it finds itself not strongly enough charged to hold together, and proceeds to emit one negative electron after another in its search for stability. Each emission transfers the nucleus to an element one step higher, without affecting its mass-number: and the authorities agree that there are at least four consecutive emissions, after the last of which the atomic number is 96! This addition of four new elements to the Periodic Table opens a new field to chemists, one which they can scarcely have expected ever to be able to enter. The four have no proper names as yet, a curious circumstance in view of the fact that discoverers of new elements have thus far been in great haste (sometimes too great haste) to name them. Mendeleieff long ago used to denote an expected but undiscovered element by prefixing "eka" to the name of the element just above the vacant place in the Periodic Table; these new four are sometimes called eka-rhenium, eka-osmium, eka-iridium and eka-platinum, but on looking at such words one is inclined to prefer the atomic numbers.

Now to summarize. The world as we knew it before the days of transmutation was constructed out of some two hundred and fifty kinds of atoms, each consisting of a nucleus surrounded by a family of electrons. Of these 250 kinds of nuclei the great majority were stable and perpetual, but some forty were unstable—doomed to perish in due time, by ejecting either alpha-particles or negative electrons.

These were the natural radioactive bodies. To these forty kinds of radioactive nuclei already found in Nature, physicists have added in a scant four years no fewer than two hundred and twenty more by the art of transmutation. Every chemical element which is known to exist at all, with the sole exception of hydrogen, has at least one radioactive type of nucleus or isotope, and many have more than one. These man-made radioactive nuclei are often made simply by adding neutrons to nuclei which already exist and are stable. There are, however, other and more complicated processes, in which neutrons or protons or deuterons or alpha-particles impinge on nuclei and seem to enter them, and other particles leap out. Many radioactive bodies have already been made in two or three different ways, some in as many as five.

Few things are riskier than to suggest a limitation either on the scope of Nature or on the possibilities of science, and many a scientist is remembered chiefly for such a suggestion which later the course of events proved foolish. Yet there are circumstances in this case which give some ground for suspecting that already we may know nearly all of the stable and may have created nearly all of the radioactive nucleus-types. Several hundreds of types have now been made by the art of transmutation, but of them nearly all which seem to be stable are not new, and nearly all which are new are radioactive. This implies that the earth has already been stocked with almost all the stable nucleus-varieties, but not necessarily that we have yet come near to making all of the possible radioactive kinds. There are however reasons for believing that most of the remaining types have so little durability, that even if they were to be made they would not last long enough to be identified as radioactive. Nature probably has come quite close to building all the imperishable forms, we possibly almost as close to creating all of those which are capable of a little but not a perpetual life. Perhaps it is fitting that people who are not immortal should not be able to construct new elements which are immortal; but we at least can rejoice in having diversified the scene of the world with a surprising number of new substances which are none the less remarkable for being transitory.

Abstracts of Technical Articles from Bell System Sources

Protection Features for the Joint Use of Wood Poles.[1] J. O'R. Coleman and A. H. Schirmer. The paper reviews the historical development of joint use and the general results to date of studies of protective problems of lower and higher voltage joint use. The safety features are reviewed from the standpoint of (1) subscribers' premises, (2) employees, and (3) telephone plant. Characteristics of equipment of power and telephone plant as far as they relate to this problem are given. The various factors which determine magnitude and duration of the current and voltage in the telephone plant resulting from a contact with power conductors are discussed. Improved methods for obtaining safety under various conditions, where higher voltage joint use is found to be the best over-all solution, are described.

High-Speed Motion Picture Photography Applied to Design of Telephone Apparatus.[2] W. Herriott. High-speed motion pictures are employed at Bell Telephone Laboratories as a visual aid in the study of problems associated with the design, manufacture, and testing of telephone apparatus. A new high-speed camera of the optical compensator type operating at 4000 pictures per second is described, and its application to the study of problems associated with telephone apparatus is discussed.

Mass Ratio of the Carbon Isotopes from the Spectrum of CN.[3] F. A. Jenkins and Dean E. Wooldridge. With a source containing carbon enriched about ten times in C^{13}, the violet CN bands have been photographed with a dispersion of 0.63A/mm. Measurements are given of the lines of low rotational quantum number in the 0,0, 0,1 and 0,2 bands of $C^{13}N^{14}$, as well as of $C^{12}N^{14}$. The vibrational constants of the normal states of both molecules are accurately determined, and give a value of the isotope mass coefficient $\rho = \omega_e{}^i/\omega_e$ of 0.97898 ± 0.00002, corresponding to a mass for C^{13} of 13.0088. This is in essential agreement with the mass-spectrograph value, and it is shown that the finer corrections to the isotope effect are negligible in this case.

[1] *Elec. Engg.*, March 1938.
[2] *Jour. S. M. P. E.*, January 1938.
[3] *Phys. Rev.*, January 15, 1938.

Composition and Structure of Hevea Latex.[4] A. R. Kemp. Data and present views relating to the composition and structure of the latex particles are presented. The number of particles in one gram of 40 per cent latex was calculated to be 7.4×10^{12} on the basis that they have an average particle mass of 0.054×10^{-12} gram (from the microscopic data of F. F. Lucas, page 146).

A study was made of the effect of several factors on the water content of rubber in pressed coagulum from fresh and treated latices. The average value for the water of retention of rubber coagula from fresh latex was found to be about 11 per cent, increasing to about 22 per cent in the case of old latex and deproteinized rubber from alkali-treated latex. This water appears to be held mechanically in the colloid hydrocarbon structure of the latex particles.

The particle structure of sheet rubber is discussed and it is suggested that plasticization by milling involves the conversion of the gel hydrocarbon shell on the rubber particles to sol rubber through oxidation.

Ultraviolet Microscopy of Hevea Rubber Latex.[5] Francis F. Lucas. Samples of bulk rubber latex received in sealed cans from two sources have been investigated by means of the ultraviolet microscope. The advantages of the ultraviolet microscope are (*a*) an enormous increase in resolving power, (*b*) selective absorption of the ultraviolet light by many substances, and (*c*) the ability to section optically very small objects suited to the purpose.

Brief descriptions of the apparatus and technic are given. Artifacts have been minimized in the preparation of the slides. A multitude of particles bordering on colloidal dimensions have been clearly resolved. Particle size measurements, including complete tabular data and a particle size distribution curve for each specimen, are given. Approximately 90 per cent of the particles are 0.50 micron or less in diameter. The shape of the latex particle appears predominantly spherical, although elongated particles and irregular shaped particles are found. Optical sections in some cases show these to be groups of particles; two particles may coalesce to form one. Many of the smaller particles appear to lose their electrical charges and become attached to larger particles. Possible effects of ultraviolet radiation are discussed.

Dielectric Losses in Polar Liquids and Solids.[6] S. O. Morgan. Dielectric loss is the energy dissipated as heat in a dielectric when it is in an electric field. Losses due to dipoles represent only one of a num-

[4] *Indus. and Engg. Chem.*, February 1938.
[5] *Indus. and Engg. Chem.*, February 1938.
[6] *Indus. and Engg. Chem.*, March 1938.

ber of possible means by which energy may be dissipated in a dielectric; there may be losses due to free ions and also due to dielectric polarizations other than dipole polarizations. However, in many materials dipoles are an important source of loss; it is the purpose of this paper to consider some of the typical cases of dipole loss and to point out some of the relations between chemical composition and dipole loss which follow from the recent experimental and theoretical study of dielectric behavior.

Order-Disorder Transformations in Alloys.[7] FOSTER C. NIX and WILLIAM SHOCKLEY. An extensive résumé of a subject which is becoming of increasing interest to physicists and metallurgists. The article divides into two parts, forty pages being devoted to the theories of the order-disorder phenomenon, and twenty pages to experimental studies of superstructures.

Thyratrons for Grid-Controlled Rectifier Service.[8] G. H. ROCKWOOD. It is common knowledge that the output voltage of a rectifier fluctuates with changes in load current and supply line voltage. Frequently these fluctuations are so large that means must be used to correct them. This is particularly true when the rectifier feeds a load having a high back electromotive force and a small resistance, such as a storage battery. The facility with which the output voltage may be controlled by the use of thyratrons as the rectifying element has encouraged the design of tubes especially suited to this purpose. There is available a variety of circuits such that the output voltage of a rectifier may be made to obey any desired law. The successful application of these circuits depends upon the degree of reliability of the thyratron tubes used in them. To be most successful the tubes must possess certain characteristics. This paper gives a brief review of the operation of grid-controlled rectifier circuits, discusses the requirements which such circuits impose on the tube characteristics, and describes a particular type of thyratron with mercury-plus-argon filling which has proved especially useful in such rectifiers.

Progress in Non-ferrous Metals and Alloys During the Past Few Years.[9] EARLE E. SCHUMACHER and ALEXANDER G. SOUDEN. The purpose of this review is to present the more important advances in the non-ferrous field during the past few years, the topics discussed being classified broadly as fundamental and practical. The former

[7] *Reviews of Modern Physics*, January 1938.
[8] *Trans. the Electrochemical Society*, Vol. LXXII, 1937, pp. 213–224.
[9] *Mining and Metallurgy—Institute of Metals Division*, January 1938.

includes those studies that have done most toward developing the basic science of metals and alloys, and the latter includes technical developments and applications.

Theory of Order for the Copper Gold Alloy System.[10] W. SHOCKLEY. The theory of order and disorder, in the form used by Bragg and Williams, is extended to arbitrary composition of the constituent elements. The work is based upon the nearest neighbor interaction assumption of Bethe and the connection between the Bethe and Bragg-Williams theory is shown. In order to extend the Bragg-Williams theory to compositions other than 25 and 50 atomic per cent, new definitions of order are developed. The results are presented in terms of phase diagrams and curves showing energy vs. temperature, specific heat vs. temperature and state order vs. temperature. These results are of importance in giving a general picture of the order-disorder transformation for a wide composition range. They are not in detailed accord with experiment due to the rather idealized picture underlying the nearest neighbor assumption.

A Theory of Noise for Electron Multipliers.[11] W. SHOCKLEY and J. R. PIERCE. The noise in secondary-emission electron multipliers is considered from a theoretical viewpoint. The noise properties of a stage are correlated with its secondary-emission properties: the mean value m and mean-square deviation δ^2 of the number of secondaries per primary. If $\overline{I_{p\Delta f}^2}$ and $\overline{I_{s\Delta f}^2}$ denote the mean-square noise current lying in the frequency band Δf in the primary- and secondary-electron currents, then $\overline{I_{s\Delta f}^2} = m^2\,\overline{I_{p\Delta f}^2} + \delta^2 2e\bar{I}_{p}\Delta f$ where $\bar{I}_p$ is primary direct current. This result is applied to many-stage multipliers. For n similar stages $\overline{I_{s\Delta f}^2} = \overline{M^2 I^2}_{p\Delta f}^2 + \delta^2[M(M-1)/m(m-1)]2e\overline{I_{p\Delta f}}$ where $M = m^n$ is the over-all gain of the multiplier.

Wave Guides for Electrical Transmission.[12] G. C. SOUTHWORTH. The transmission of electric power at extremely high frequencies through rods or "wires" of dielectric and through metal tubes, without the usual return conductor, was predicted mathematically many years ago. Recently experiments have confirmed this theory. Wave guides offer the possibility of transmitting very wide frequency bands and consequently extremely large numbers of speech channels without the high attenuations encountered in radio; in fact, constantly decreasing attenuation with increasing frequency is predicted for one type of wave.

[10] *Jour. Chemical Physics*, March 1938.
[11] *Proc. I. R. E.*, March 1938.
[12] *Elec. Engg.*, March 1938.

Some of the properties of the waves and the apparatus used in studying them are described in this article.

Recent Development in Hill and Dale Recorders.[13] L. VIETH and C. F. WIEBUSCH. A new sound-on-disk recorder has been developed in which is used the principle of feeding part of the output of the system back to the input of the associated driving amplifier in properly controlled relationship. The use of this principle, which is widely used in feedback amplifiers, replaces the usual practice of providing dissipative elements for the control of an electrically driven vibrating system. Heretofore no practical application of feedback to electromechanical systems has been made, possibly because the requirements for stable operation of such systems are difficult of achievement. Through recent developments these requirements have been satisfactorily met. The new recorder is capable of recording on wax or direct-recording material without appreciable effect upon its characteristics, which include uniform response from 30 to 12,000 cps. and exceptional freedom from distortion. The recorder is extremely simple and affords easy means for field calibration from the feedback element, whose output is in direct proportion to the stylus velocity. These means also make available a monitoring voltage which, properly amplified, gives a precise aural picture of the stylus behavior during recording.

Internal Friction in Solids—III. Experimental Demonstration of Thermoelastic Internal Friction.[14] C. ZENER, W. OTIS and R. NUCKOLLS. In order to demonstrate the presence of thermoelastic internal friction, the authors measured the internal friction of a copper reed over a wide frequency range (50 to 4000 cycles/sec.). They obtained a maximum precisely at the predicted frequency. The observed variation of internal friction with frequency proves that, over a wide frequency range, the internal friction due to the flow of heat back and forth across a reed is of a larger order of magnitude than that due to all other causes. Independent experiments of Bennewitz and Rötger on wires of silver, aluminum, brass, steel, and glass are shown to furnish an equally striking demonstration of thermoelastic internal friction.

[13] *Jour. S. M. P. E.*, January 1938.
[14] *Phys. Rev.*, January 1, 1938.

Contributors to this Issue

H. W. Bode, A. B., Ohio State University, 1924; M.A., 1926; Ph.D., Columbia University, 1935. Bell Telephone Laboratories, 1926–. Dr. Bode has been engaged in the study of transmission networks, such as wave filters, attenuation equalizers, and phase correctors.

James A. Carr, B.S. in Electrical Engineering, Virginia Polytechnic Institute, 1919; Instructor in Electrical Engineering at Massachusetts Institute of Technology, 1920. American Telephone and Telegraph Company, Development and Research Department, 1921–27; Bell Telephone Laboratories, 1927–. Mr. Carr has been engaged principally in outside plant development work.

Karl K. Darrow, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917. Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences.

C. O. Gibbon, B.S., Purdue University, 1914; M.S., Massachusetts Institute of Technology, 1917; M.E.E., Harvard University, 1917; Electrical Engineering Laboratory Assistant, Massachusetts Institute of Technology and Harvard University, 1917–18. American Telephone and Telegraph Company, Department of Operation and Engineering, 1921–. First engaged in work on telegraph matters, Mr. Gibbon has since 1925 been occupied principally with problems pertaining to local transmission and in the preparation of engineering data for the transmission design of the exchange plant.

F. V. Haskell, B.S. in Electrical Engineering, Worcester Polytechnic Institute, 1926. New York Edison Company, 1926–29. Bell Telephone Laboratories, 1929–. Mr. Haskell has been engaged principally in aerial systems work.

Arthur W. Horton, Jr., A.B., Princeton University, 1920; E.E., Princeton University, 1922. Western Electric Company, Engineering Department, 1922–25; Bell Telephone Laboratories, 1925–. Mr. Horton has been engaged in work relating to the application of voice operated devices in the toll telephone plant and in studies of toll transmission.

O. J. Murphy, B.S. in Electrical Engineering, University of Texas, 1927; Columbia University, 1928–31. Bell Telephone Laboratories, 1927–. Mr. Murphy has been principally engaged in studies of the effects of transmission delay and voice operated devices on toll telephone circuits.

A. C. Norwine, A.B., University of Missouri, 1923; B.S. in Electrical Engineering, 1924; E.E., 1925. Bell Telephone Laboratories, 1925–. Mr. Norwine has been principally engaged in studies of the effects of transmission delay and voice operated devices on toll telephone circuits.

VOLUME XVII JULY, 1938 NUMBER 3

THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

Hertz, the Discoverer of Electric Waves—*Julian Blanchard* . 327

Instruments for the New Telephone Sets—*W. C. Jones* . . 338

Transmission Features of the New Telephone Sets —*A. H. Inglis* 358

Spectrochemical Analysis in Communication Research —*Beverly L. Clarke and A. E. Ruehle* 381

High Speed Motion Picture Photography—*W. Herriott* . . . 393

An Optical Harmonic Analyzer—*H. C. Montgomery* 406

Magnetic Shielding of Transformers at Audio Frequencies —*W. G. Gustafson* 416

Coaxial Cable System for Television Transmission —*M. E. Strieby* 438

Stabilized Feedback Oscillators—*G. H. Stevenson* 458

The Discovery of Electron Waves—*C. J. Davisson* 475

Abstracts of Technical Papers 483

Contributors to this Issue 486

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

EDITORIAL BOARD

F. B. Jewett	H. P. Charlesworth	W. H. Harrison
A. F. Dixon	O. E. Buckley	O. B. Blackwell
D. Levinger	M. J. Kelly	H. S. Osborne
	W. Wilson	

R. W. King, *Editor* J. O. Perrine, *Associate Editor*

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are fifty cents each. The foreign postage is 35 cents per year or 9 cents per copy.

Copyright, 1938
American Telephone and Telegraph Company

PRINTED IN U. S. A.

HEINRICH RUDOLPH HERTZ

1857–1894

The Bell System Technical Journal

Vol. XVII *July, 1938* *No. 3*

Hertz, the Discoverer of Electric Waves *

By JULIAN BLANCHARD

FIFTY years have passed since those memorable researches of the young German physicist, Heinrich Rudolph Hertz, which have come to be regarded as the starting point of radio. For it was he who first detected, and measured, electromagnetic waves in space—waves which had been predicted, it is true, but which had never before been observed. It is not to be claimed, of course, that the radio art would have failed to be born were it not for his genius, for we know that almost simultaneously the experiments of Lodge in England were pointing with certainty to the same discoveries, and the speculations of others were revolving around the possibility of generating electric waves. Yet it was the remarkably clear vision of Hertz, combined with his consummate persistence and skill, that won for him the prize and justly enshrined his name among the immortal men of science.

So, upon this golden anniversary of the opening of a new epoch in the realm of communication, it is fitting that we pause to do honor to his memory and to consider anew the significance of his great accomplishment.

The formal facts of Hertz's biography can be set down very quickly. He was born at Hamburg on February 22, 1857, his father an attorney, belonging to a family of successful merchants, his mother the daughter of a doctor of medicine, and the descendant of a long line of Lutheran ministers—all of cultural tastes and attainments on both sides. At the age of twenty he went off to school at Munich, after a rather unorthodox preparatory training, supposedly to pursue an engineering career, but he was torn between this resolve and his natural inclination for the study of pure science. Soon after reaching Munich he felt compelled to put the matter before his parents, to whom he frequently and confidingly wrote concerning his plans and his work. In a long letter written in November, 1877, he said, "I really feel ashamed to say it, but I must: now at the last moment I want to change all my

* Published in *Proc. I.R.E.*, May, 1938, Vol. 26, No. 5.

plans and return to the study of natural science. I feel that the time has come for me to decide either to devote myself to this entirely or else to say good-bye to it; for if I give up too much time to science in the future it will end in neglecting my professional studies and becoming a second-rate engineer. Only recently, in arranging my plan of studies, have I clearly seen this—so clearly that I can no longer feel any doubt about it. . . ." And then follows a lofty and appealing presentation of the reasons for his choice. Concluding, he wrote: "And so I ask you, dear father, for your decision rather than for your advice; for it isn't advice that I need, and there is scarcely time for it now. If you will allow me to study natural science I shall take it as a great kindness on your part, and whatever diligence and love can do in the matter, that they shall do. I believe this will be your decision, for you have never put a stone in my path, and I think you have often looked with pleasure on my scientific studies . . ."

Matters were arranged as he wished and he joyfully pursued his studies at the University and at the Technical Institute, working hard on mathematics and mechanics, and spending much time in the physical laboratory. In October, 1878, he went to Berlin and became there a student under the mighty giants, von Helmholtz and Kirchhoff; writing to his parents in November, "I am now thoroughly happy, and could not wish things better." In 1880 he gained his doctorate, and in October of that year was appointed assistant to Helmholtz, which delightful and stimulating association continued until Easter of 1883. He then went to the University of Kiel to become lecturer in theoretical physics, and here he first began to reflect seriously upon Maxwell's electromagnetic theory of light.

He was soon promoted again, and at Easter, 1885, he became professor of experimental physics at the Technical High School in Karlsruhe. Here, in 1886, at the age of twenty-nine, he married Elizabeth Doll, daughter of the professor of geodesy in the same institution, and their home became a congenial meeting place for their many cultured associates. It was at Karlsruhe also that he began his researches on electric waves. Before they were finished he was called in 1889 to succeed the celebrated Clausius at the University of Bonn, thus at the age of thirty-two having arrived at a position in the academic world not ordinarily to be achieved until much later in life.

He soon thereafter relinquished to others the further exploration of the great new territory of electric waves he had opened up and returned to some investigations on the discharge of electricity in rarefied gases, a subject which had interested him while at Berlin. He then devoted his attention entirely to what proved to be his last work, a

treatise on mechanics. In the summer of 1892 he suffered a severe illness which eventually led to chronic blood poisoning, of which he died, after indescribable suffering, on January 1, 1894. He would have been just thirty-seven years old the following month.

Although the fame of Hertz rests primarily on his electric wave researches, these constitute by no means the whole of his work. His collected papers, edited by the German physicist, Dr. Philipp Lenard, and admirably translated into English by Professor D. E. Jones and associates, comprise three volumes. The first consists mainly of his miscellaneous earlier papers, some twenty-odd titles altogether. One of these, published in 1884, "On the Relations between Maxwell's Fundamental Electromagnetic Equations and the Fundamental Equations of the Opposing Electromagnetics," marked an important step in the development of Hertz's ideas, and has been called his greatest contribution to theoretical physics. In it he opposed the old orthodox theories of electric phenomena based on action at a distance, which were supported by most of the Continental physicists, and definitely aligned himself with the followers of Maxwell. This volume also contains his semipopular Heidelberg lecture of 1889, "On the Relations between Light and Electricity," giving a general account of his more recent work. It strikingly illustrates the charm and felicity of style he could employ in the presentation of a difficult subject, and cannot fail to be read, and reread, with pleasure and admiration. The volume ends with a eulogy of Helmholtz on the occasion of his seventieth birthday, wherein the pupil, in equally graceful language, paid homage to his beloved and inspiring preceptor. These two papers reveal so clearly between the lines the manner of man their author was.

The second volume contains the papers on electric waves, which had been collected by the author himself, as a result of numerous requests for reprints, and published under the German title, "Untersuchungen über die Ausbreitung der Elektrischen Kraft," and later in English as "Electric Waves," with a lucid introduction by Hertz explaining the motive and significance of each of the separate papers. One of the first of these describes an important by-product of the electric wave studies, the discovery of the effect of ultraviolet light upon an electrical discharge. This discovery itself uncovered a new wealth of physical problems and the subject became immediately of great interest to many experimenters, most of whom at the time little suspected that this subsidiary effect was not the main discovery, or imagined that electrical science was on the eve of much greater conquests. The paper attracted attention to Hertz and aroused a popular nterest in him, so that everything coming from him thereafter was

eagerly read. Had it not been for this rather accidental circumstance the importance of his subsequent work might have gone temporarily unnoticed, as has happened with some of the greatest discoveries.

The third volume in the series is his book, "Die Prinzipien der Mechanik," completed with difficulty during his last illness and published a few months after his death. A lengthy preface by the venerable Helmholtz gives an appreciative sketch of the author's life and work—this time a sorrowful tribute from master to pupil.

In order to understand and appraise the work of Hertz on electric waves, it will be needful to review briefly the ideas about light and electricity that prevailed at his time and before. With regard to light, Newton's corpuscular theory had given way in the early 1800's to the wave theory of Young and Fresnel, and long before Hertz's day the idea of transverse vibrations in a hypothetical elastic-solid medium called the luminiferous ether had become firmly established. Length of wave and velocity had been measured many times. The wave theory accounted satisfactorily for all the known phenomena in optics and there was no doubt in anybody's mind about its essential correctness, regardless of any difficulties encountered in explaining the nature of the ether and its relation to matter.

As to electricity and magnetism, the older theories of instantaneous action at a distance were beginning to weaken with the discoveries, in the early part of the nineteenth century, of the reactions between electric currents and magnets, and the phenomenon of induction. Hitherto there had been no postulation of an intervening medium to explain the transmission of the force between two charged bodies, and it was supposed that electricity and magnetism, like gravitation, acted across empty space in straight lines and instantaneously. In somewhat different dress these ideas were given new life around the middle of the century, particularly by some of the German physicists. To Faraday, however, such views were unacceptable. He wished to get rid of the idea of action at a distance and in his mind pictured a medium, along the contiguous molecules or particles of which the force was propagated. In this medium he visualized "lines of force" emanating from or terminating upon the electric charges or magnetic poles, acting like stretched elastic threads, repelling each other sidewise as well as tending to contract. Thus, in his thinking, attention was focused upon the insulating medium surrounding a conductor, the "dielectric" as he called it, for here, he thought, was the real seat of the action. He believed also that there existed some direct relation between light and electricity or magnetism. He was ever seeking to find such a relation and in the course of his many experiments he discovered the rotation

of the plane of polarization of light by a magnetic field. In his speculations there was one question which was continually presenting itself to him. Do electric and magnetic forces, like light, require time for their propagation? Are there waves? But to this question he was unable to find an answer.

And then came Maxwell, building upon the foundation that Faraday had laid, translating Faraday's ideas into the language of mathematics, and making the grand generalization that light and electric waves are one and the same phenomenon, propagated by the same medium, with the same velocity, and differing only in wave-length. Like Faraday, he considered the energy of the electromagnetic field to reside in the dielectric. He conceived the medium to have properties analogous to those of an elastic solid, which would spring back to its original state upon the removal of the straining force. The alteration of the displacement, or "polarization," in the medium was viewed by him as an electric current, which he called the "displacement current," as distinguished from the "conduction current" existing in conductors. From the general equations which he formulated it was shown that only transverse vibrations (like light) could be propagated in such a medium and that the velocity of propagation should be equal to the ratio of the two systems of electrical units, that is, to the number of electrostatic units of electricity contained in one electromagnetic unit. This ratio had been experimentally determined by Weber and Kohlrausch (it was later redetermined by Maxwell himself, by a different method), and the fact that it agreed so very closely with the measured velocity of light was one of the strongest points in favor of the view that light waves were identical with the hypothetical electromagnetic waves. Maxwell's comprehensive theory was first enunciated in his 1865 paper entitled "A Dynamical Theory of the Electromagnetic Field," and afterwards elaborated in his great "Treatise on Electricity and Magnetism," published in 1873, but for a number of years it was regarded by many as merely a speculation, by others as probably true, and by none as conclusively proved. It remained for Hertz to add the capstone to the theory by actually demonstrating for the first time the existence of electromagnetic waves in space; and furthermore, to show experimentally that they had all the physical properties of ordinary light waves.

In 1888, while he was teaching at the Technical High School in Karlsruhe, Hertz carried out the brilliant experiments which have made his name famous. These were actually a part of a long series of experiments which began in 1886, and which came about partly by accident, and yet were the result of his keen interest in everything

connected with electric oscillations; an interest extending back to 1879, when, at the suggestion of Helmholtz, he had considered tackling a prize problem proposed by the Berlin Academy of Science aimed at the proof of a portion of Maxwell's theory, but which he had abandoned for the reason that oscillations of sufficiently high frequency were not then available. While using in his lectures at Karlsruhe a pair of flat ("pancake") coils, called Reiss or Knockenhauer spirals, mounted adjacent to each other, he was surprised to find how easy it was to obtain sparks between the terminals of the secondary coil when a small Leyden jar or even a small induction coil was discharged through the primary, provided the primary discharge took place across a spark gap. This, of course, was an indication of an exceptionally strong inductive effect. This observation led to his discovery of a method of exciting electric disturbances far more rapid than any hitherto known, such as those of Leyden jars or open induction coils as customarily used, and having wave-lengths, it turned out, capable of being measured within the confines of a laboratory. His oscillator was nothing more than a short metal rod with a spark gap in the middle (sometimes with metal spheres or plates attached to the ends, resembling a dumbbell), the sparking terminals consisting of small knobs or spheres which were connected to the terminals of a Ruhmkorff induction coil; the small inductance and capacitance of this simple linear conductor, together with the proper functioning of the spark gap, accounting for its success. By such means Hertz obtained wave-lengths from a few meters down to 30 centimeters, and so began, it is seen, with the "ultrashort" waves that are again coming into vogue.

Hertz began his experiments with a study of the "induction" about this exciter. As he commented in his first paper in this series, "On Very Rapid Oscillations," published in May, 1887, theory had predicted the possibility of very rapid oscillations in open-wire circuits of small capacitance, but it could not be predicted from theory whether they could be produced on such a scale as to admit of their being observed. Hertz not only devised a method of producing such oscillations, but also discovered a method of detecting them, by their effects in the surrounding space. His detector consisted merely of a short length of wire bent in the form of a rectangle or a circle and containing a micrometer spark gap, across which minute sparks could be seen in a darkened room; especially if this secondary circuit was in electrical resonance with the exciter. This exceedingly simple detector was indeed a capital discovery. Some five years earlier Professor G. F. Fitzgerald, of Dublin, had suggested "the combination of a vibrating generating circuit with a resonant receiving circuit . . . as one by

which this very question might be studied." But, as he said afterwards in speaking of Hertz's work, "I did not see any feasible way of detecting the induced resonance: I did not anticipate that it could produce sparks." Concerning this contrivance the following interesting remarks were made by its author in his Heidelberg lecture above referred to: "The method had to be found by experience, for no amount of thought could well have enabled one to predict that it would work satisfactorily. For the sparks are microscopically short, scarcely a hundredth of a millimeter long; they only last about a millionth of a second. It seems absurd and almost impossible that they should be visible; but in a perfectly darkened room they *are* visible to an eye which has been rested in the dark. Upon this thin thread hangs the success of our undertaking." *Multum in parvo*, truly!

After a series of preliminary experiments, in which he studied the various induction effects, including the phenomenon of resonance, and demonstrated waves on wires (an earlier, but overlooked, discovery of von Bezold, in 1870), and also solved the problem of the Berlin Academy—"to establish experimentally any relation between electromagnetic forces and the dielectric polarization of insulators"—he was fully convinced that the disturbance was propagated through space, independently of wires, with a finite velocity and in the form of waves, in accordance with Maxwell's prediction. His conclusion was then definitely and convincingly proved by making use of the well-known method of reflection and interference to produce standing waves, and noting the position of the nodes and antinodes. These epochal experiments were described in a paper entitled "Electromagnetic Waves in Air and Their Reflection," published in May, 1888. But he did not stop there, and in these and succeeding investigations he showed that electric waves are reflected from plane and curved metal surfaces in accordance with the same laws as light waves; that they are refracted in passing thorugh prisms of pitch, paraffin, and other dielectrics; and that they are polarized by a grating of parallel wires, and hence are transverse waves. From actual measurements of their wave-length and computations of their frequency (from the constants of his oscillator), he calculated their velocity and found that it was the same as the velocity of light. As summarized by Hertz himself, "The object of these experiments was to test the fundamental hypotheses of the Faraday-Maxwell theory, and the result of the experiments is to confirm the fundamental hypotheses of the theory." The old action-at-a-distance philosophy had come to an end.

The importance of Hertz's contributions to this great subject received instant and enthusiastic recognition, and his experiments were

soon being repeated in all the important laboratories of the world. The English mathematical physicist, Oliver Heaviside, writing in 1891, said: "Three years ago electromagnetic waves were nowhere. Shortly after, they were everywhere." Here were researches of a most abstruse and complex character, with no apparent utility and having no elements of popular appeal, and yet bringing to their author such acclaim as had seldom been accorded to a man of science. Honors were showered upon him on every hand, at home and abroad. In England, where his work was especially appreciated, he was awarded the coveted Rumford medal by the Royal Society.

Hertz's characteristic modesty in referring to his own achievements was matched only by his generosity in giving credit to the accomplishments of others. In one of his lectures he said, "Such researches as I have made upon this subject form but a link in a long chain. . . . Lack of time compels me, against my will, to pass by the researches made by many other investigators; so that I am not able to show you in how many ways the path was prepared for my experiments, and how near several investigators came to performing these experiments themselves." Mention has been made of the investigations of Sir Oliver Lodge in the same field and the imminence of his discovery of the same phenomena. It is pleasant, indeed, in this instance to be able to record the absolute lack of any feeling of jealousy or envy on the part of either of these courteous gentlemen. In the introduction to his collected papers Hertz wrote, "I may here be permitted to record the good work done by two English colleagues who at the same time as myself were striving towards the same end. In the same year in which I carried out the above research, Professor Oliver Lodge, in Liverpool, investigated the theory of the lightning conductor, and in connection with this carried out a series of experiments on the discharge of small condensers which led him on to the observation of oscillations and waves in wires. Inasmuch as he entirely accepted Maxwell's views and eagerly strove to verify them, there can scarcely be any doubt that if I had not anticipated him he would have succeeded in observing waves in air, and thus also in proving the propagation with time of electric force. Professor Fitzgerald, in Dublin, had some years before endeavored to predict, with the aid of theory, the possibility of such waves, and to discover the conditions for producing them."

On his part Lodge just as generously wrote, only a few years afterwards, in an obituary of his rival: "Hertz stepped in before the English physicists, and brilliantly carried off the prize. He was naturally and unaffectedly pleased with the reception of his discovery in England, and his speech on the occasion of the bestowal of the Rumford medal

by the Royal Society will long be remembered by those who heard it for its simplehearted enthusiasm and good-feeling. His letters are full of the same sentiment. . . ."

Noteworthy, indeed, was the extreme modesty of this scientific lion of the hour, and equally striking his consideration for the feelings of others. It is recorded that when the Royal Society presented him with the Rumford medal, "he silently disappeared from Bonn for a few days —none knew why—and he returned as silently." In refusing the request made by the editor of *The Electrician* (of London) in 1890 for his photograph, Hertz replied, "I feel as if presenting my portrait now in so prominent a place follows too quickly the little work I have done. I should like to wait a little, and see if the general approbation which my work meets with is of a lasting kind. Too much honor certainly does me harm in the eyes of reasonable men, as I have sometimes occasion to observe. If your kind intention is the same in two years, even one year, I shall readily consent and help you in every respect." Four years later the portrait was published, following upon Hertz's death.

Upon the untimely ending of his brief but brilliant career, occurring in the very prime of life, before he was yet thirty-seven, there was a feeling of shock and sadness in every scientific quarter. Many were the sincere tributes paid to his memory, honoring him for his rare personal qualities as well as his distinguished scientific attainments. Some expressions from Lodge have already been quoted. Said he in his obituary in *The Electrician*, "Not a student of physical science on the planet but will realize and lament the sad loss conveyed by the message, 'Hertz is dead.'" The editor of that journal wrote, "In the modesty and self-forgetfulness which blend so admirably in the spirit of true scientific research Hertz was singularly rich." In an editorial note in an American journal, *The Physical Review*, we find the following: "In addition to the recognition of those who were able to appreciate his work, Hertz received the acclamations of the entire world of thought. Fortunately he possessed a nature of such complete simple-mindedness that his sudden rise into a position akin to notoriety had no effect upon him. The unassuming bearing which had always characterized him remained with him to the end."

In a memorial address delivered by Professor Herman Ebert before the Physical Society of Erlangen on March 7, 1894, the following sentiments are expressed: "In him there passed away not only a man of great learning, but also a noble man, who had the singular good fortune to find many admirers, but none to hate or envy him; those who came into personal contact with him were struck by his modesty and charmed by his amiability. He was a true friend to his friends, a respected teacher

to his students, who had begun to gather round him in somewhat large numbers, some of them coming from great distances; and to his family he was a loving husband and father."

It can be said in retrospect that the fundamental invention in radiotelegraphy was made by Hertz, and yet it is true that the discoverer of electric waves had no anticipations as to their utilitarian possibilities. There was no rush to the patent office; indeed, it was not until two years after Hertz's death that the first application for a radio patent was filed, by Marconi. The chief interest at the time was purely scientific, the results being hailed as the settlement of a great scientific controversy, the confirmation of Maxwell's theory, the annexation to electricity of the entire domain of light and radiant heat. In the current literature we find little of prophecy with respect to utility. Sir William Crookes has been credited with being one of the first to foresee distinctly the applicability of "Hertzian" waves to practical telegraphy. In an article in the *Fortnightly Review* for February, 1892, he made a remarkably accurate forecast of what was to come: "simpler and more certain means of generating electrical rays of any desired wave-length"; "more delicate receivers which will respond to wave-lengths between certain defined limits and be silent to all others"; "means of darting the sheaf of rays in any desired direction. . . ." And for secrecy he foresaw that "the rays could be concentrated with more or less exactness on the receiver," if the sender and receiver were stationary; or, if moving about, "the correspondents must attune their instruments to a definite wave-length. . . ." "This is no mere dream of a visionary philosopher," he wrote. "All the requisites needed to bring it within the grasp of daily life are well within the possibilities of discovery, and are so reasonable and so clearly in the path of researches which are now being actively prosecuted in every capital of Europe that we may any day expect to hear that they have emerged from the realms of speculation into those of sober fact."

As we well know, all that he predicted, and more, has become reality, although progress was not to be as rapid as then seemed probable. One of those who at the time had the imagination to see, if only hazily perhaps, the great possibilities of Hertz's discovery was the youthful Marconi, who had also the initiative and the determination to put his ideas into execution, to make the new-found waves useful to mankind. Within a few years, around the turn of the new century, the world was to be thrilled by the detection of a wireless signal transmitted across the wide Atlantic. But there were insurmountable limitations to the means at hand, and it remained for still another wave in the onward roll of science, the advent of the magical era of electronics, to yield the

tools necessary for the really great advance that was ahead. With the invention and development of the amplifying and oscillating vacuum tube progress was greatly accelerated, and in a comparatively short time there had been achieved, by a veritable army of experimenters, the marvels of world-wide intercommunication which are so familiar to us today.

Instruments for the New Telephone Sets*

By W. C. JONES

Transmitters and receivers for use at subscribers' telephone stations have been designed which not only materially improve transmission but also simplify manufacture and facilitate maintenance. This paper discusses these improvements and describes some of the new design technique employed in their development.

AS a result of continuous development work on transmitters and receivers for use at subscribers' telephone stations, new instruments have been designed which not only materially improve transmission but also embody features which simplify manufacture and facilitate maintenance. These instruments are now being produced for use in handsets, desk stands and wall sets in the Bell System.[1]

In many respects these instruments represent outstanding advances in transmission instrument design and performance. It is the purpose of this paper to discuss these improvements and to describe some of the new design technique employed in their development. The data presented will be confined almost entirely to physical measurements which serve to define the performance characteristics of the instruments. The interpretation of these data in terms of their relationship to the characteristics of associated apparatus and their overall reaction on transmission in the telephone plant is covered by a companion paper dealing with the transmission features of the new sets.[2]

Handset Applications

The new transmitter unit with an adapter was first introduced in 1934 as a replacement for the earlier type of handset transmitter.[3] There are now about five million of these transmitters in use in the plant of the Bell System. While experience has shown that they effect an outstanding improvement in performance they do not take full advantage of the possibilities of the unit type of construction from the standpoint of simplification, owing to the fact that a number of additional parts are required to mount the unit on the existing type of handset handle. The advantages of the unit type of instrument have been realized in a new design of handset introduced during 1937, about a million of which have been produced. This handset is shown

* Presented at A.I.E.E. Summer Convention, Washington, D. C., June 21, 1938.

with the new combined set in the photograph, Fig. 1, and in cross-section on Fig. 2.

In designing this handset every effort has been made to obtain the maximum degree of simplicity consistent with the electrical requirements involved and at the same time to secure an attractive design which harmonizes with the other station apparatus on the subscriber's premises. Only three phenol plastic parts are employed; namely, the

Fig. 1—Handset and desk stand equipped with the new instruments.

handle and the transmitter and receiver caps. In designing these parts particular attention has been paid to providing adequate cross-sections at the points of maximum stress and to distributing the weight so as to reduce to a minimum the breaking moments which are developed when the handset is dropped. The transmitter and receiver caps serve the dual purpose of holding the units in place and providing mechanical protection. In addition they thoroughly insulate the user from all the metal parts which are included in the electrical circuit. Both caps have smooth surfaces which can be readily cleaned. As will be pointed out later, the grid of the receiver cap also has a transmission function and plays an important part in determining the response in the upper frequency range. Spring contacts are provided to facilitate the assembly of the units in the handle. This operation is further facilitated by the fact that specific alignment of the units

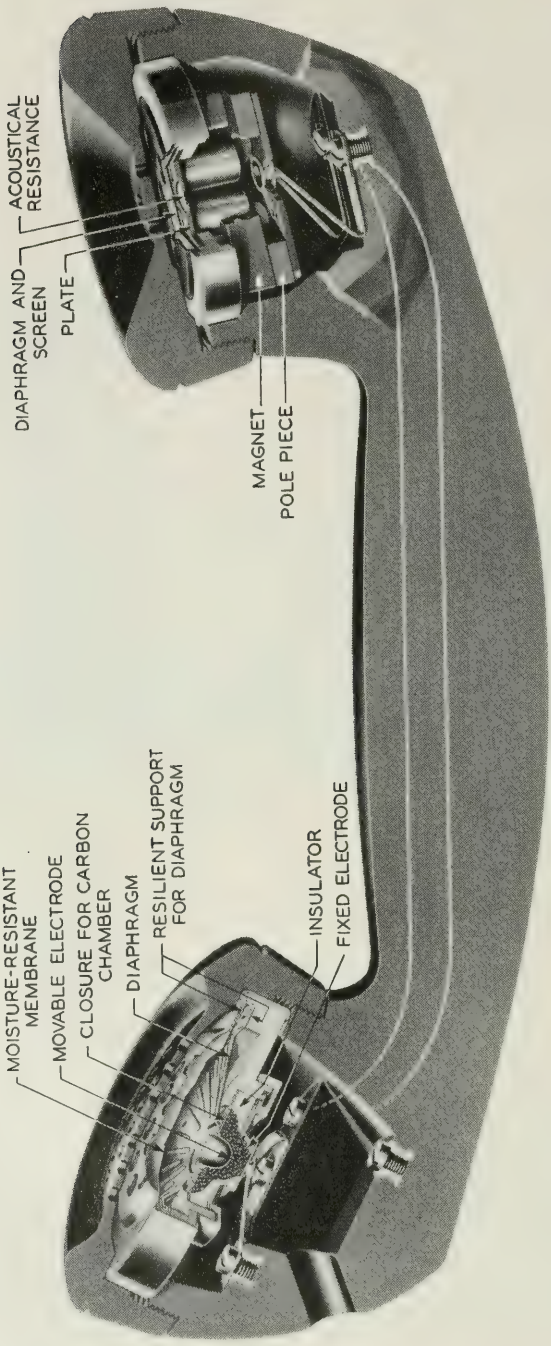


Fig. 2—Cross-section of the handset.

and the caps relative to the handle is unnecessary. The spacing between the transmitter and receiver is such that the handset can be used with the existing type of desk mounting as well as with the new combined set.

Desk Stand and Wall Set Applications

The photograph, Fig. 1, also shows the new transmitter and receiver unit adapted to desk stand and wall set use. Cross-sections of these instruments are shown on Fig. 3. The faceplate, mouthpiece and pro-

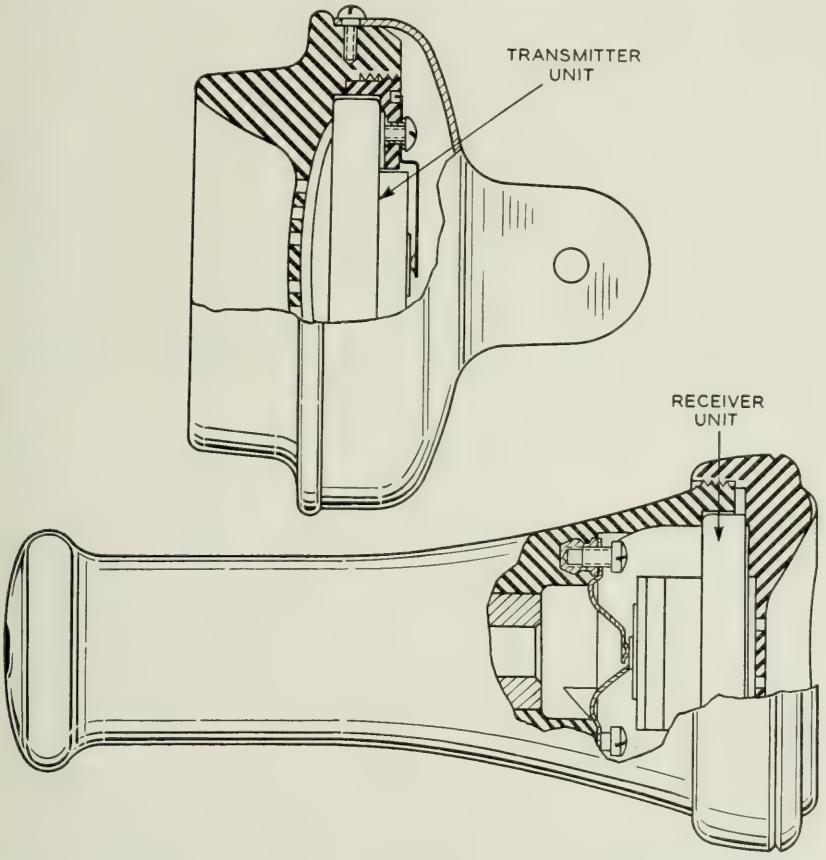


Fig. 3—Cross-sections of the transmitter and receiver for desk stands and wall sets.

tective grid of the transmitter are combined in one phenol plastic part which is so designed as to reduce cavity resonance to a minimum and provide response characteristics essentially the same as those of the handset transmitter. On the other hand, the mouthpiece is sufficiently

prominent to encourage the user to talk directly into it and in this way reduce the losses which often result when flush type faceplates are employed with desk stand and wall set instruments. A phenol plastic part, equipped with contact springs, holds the unit tightly in the faceplate and provides electrical connections.

As in the handset the unit of the receiver is held in place by the cap. Springs are provided in the shell for bringing out the electrical connections. A metal insert adds sufficient weight to meet the switchhook requirements of the existing sets. The phenol plastic parts of both the receiver and transmitter are so designed as to insulate thoroughly the units and minimize breakage.

Transmitter Unit

The new transmitter unit is of the "direct action" type, that is, one in which the movable electrode serves the dual purpose of contact and pressure surface. As is shown by Fig. 2, this electrode is mounted at the center of a diaphragm of thin aluminum alloy formed into a shallow cone and ribbed to add rigidity. "Books" of thin impregnated paper mounted in a recess in a die-cast frame provide a resilient support for the edge of the diaphragm. The fixed electrode is held in place in the frame by a threaded ring and is insulated from the frame by a phenol fibre washer and a ceramic insulator which also forms one of the surfaces of the carbon chamber. The active surfaces of both electrodes are gold plated. A silk annulus clamped at its outer edge between the ceramic insulator and the frame and its inner edge between the movable electrode and the diaphragm forms a resilient closure for the carbon chamber. Electrical connection between the movable electrode and the frame is provided by means of metal strips of low stiffness. Provision is made for machine filling the carbon chamber through a hole in the fixed electrode and closing this hole by means of a cap which crimps over a projecting shoulder. The exposed surfaces of the cap and the threaded ring are silver plated and form the contact surfaces for the electrical connections. A moisture-resistant membrane protects the internal parts of the unit from the effects of condensed moisture from the breath. This membrane is clamped at its outer edge between a protective grid of perforated metal and the frame. A thin metal ferrule fastens the grid to the frame. The exposed parts of the unit are anodically finished to resist corrosion.

In addition to being simpler than the earlier transmitter and hence less difficult to produce, the new transmitter unit has characteristics such that:

1. Its performance is less affected by angular position.
2. There is less aging under the conditions encountered in service.
3. The electrical output is higher and the response more uniform.
4. The modulation products resulting from non-linearity are materially reduced.

Effect of Angular Position.—In order to insure good contact between the carbon granules and the diaphragm in the positions in which the handset is most likely to be held in service, the carbon chamber of the earlier transmitter was placed in front instead of the conventional location in back of the diaphragm.[3] The positional characteristics of this transmitter were further improved by the use of a "barrier" type of variable resistance element in which the electrodes are stationary and form the walls of the carbon chamber, and in which the surface of the diaphragm in contact with the granules is insulated and serves only as means for changing the contact forces between the granules in response to the variations in sound pressure at the diaphragm surface. While this transmitter represented a distinct advance in handset performance from a transmission standpoint and was quite effective in reducing undesirable positional effects, particularly in the "horizontal face-up" position, it was somewhat complicated mechanically and involved the problem of providing a closure between the diaphragm and the adjacent electrode which would be sufficiently resilient to meet the transmission requirements and at the same time prevent carbon leakage. In addition, there was some degradation in quality when it was held in the "horizontal face-down" position where the carbon granules tended to fall away from the diaphragm. While this condition occurred only infrequently in service, it was one which it was considered desirable to eliminate if this could be accomplished without making the structure mechanically complex or difficult to manufacture or maintain. A tendency also was observed in the field for the resistance to increase sufficiently under certain conditions to react adversely on the operation of the associated signaling apparatus. Owing to the inherently small areas of the sound passages leading to the diaphragm the moisture condensed from the breath could not be excluded by a membrane without complicating the structure and adding sufficient mechanical impedance to impair transmission.

Following the introductory work on the barrier transmitter, an intensive study of the direct action type of carbon element was made to determine whether the limitations of the earlier structures of this type, which arose from the non-fluid character of the carbon, could be overcome. This study resulted in the transmitter unit shown on Fig. 2. This unit eliminates the undesirable features of the inverted type without sacrificing its desirable characteristics.

The electrode surfaces of the new transmitter unit are so proportioned and so spaced relative to each other that the important current paths shift their locations in the carbon mass with changes in angular position in a manner such that the mean effective pressures in the paths and the lengths of the paths result in substantially constant resistance in all positions. Furthermore, the components of the axial motion of the diaphragm effective in changing the contact forces in the paths are also such as to produce essentially constant modulation. Not only is the total resistance of the paths between the electrodes substantially

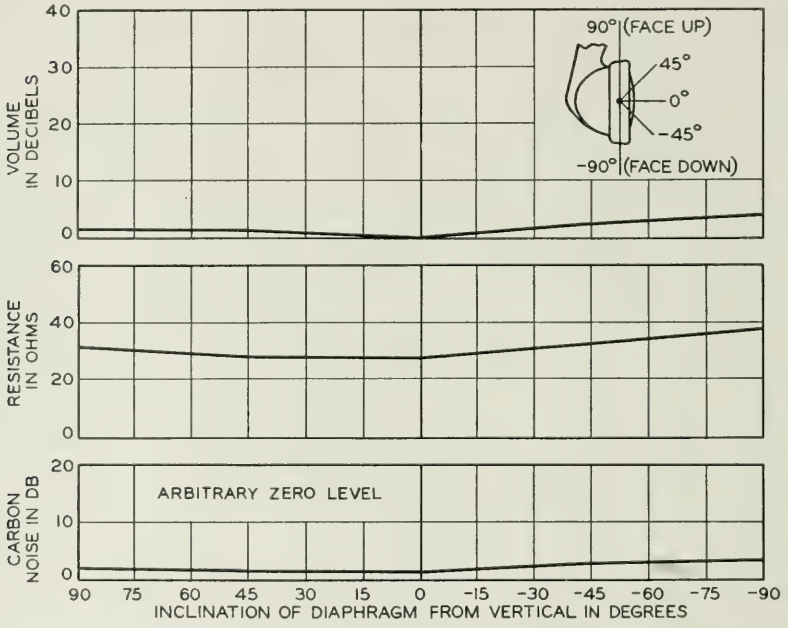


Fig. 4—Positional characteristics of the transmitter.

constant, but this resistance also is uniformly distributed between the individual contacts with the result that at no time does the contact potential rise to a value sufficiently high to produce objectionable carbon noise or "burning." These features result in resistance, volume efficiency and carbon noise characteristics which, as is shown by Fig. 4, are essentially independent of angular position.

Another and perhaps a more exacting criterion of the adequacy of a transmitter from the standpoint of its ability to function satisfactorily in all positions is the extent to which its transmission characteristics at normal speech intensities are adversely affected when immediately

preceded by loud speech. If a poorly designed transmitter is held in a position such that the carbon granules tend to fall away from the movable electrode when this test is applied, the non-fluid action of the carbon will prevent the reestablishment of contact with the electrode surface with the result that volume losses of as much as 20 db and a serious degradation in quality take place. Furthermore, these losses persist until the transmitter is jarred or moved about sufficiently to change the configuration of the granules. On the other hand, if the effect of the frictional forces within the granular mass has been taken fully into account in the design of the carbon element, these forces will not react in a manner such as to prevent good contact with the electrode following the large amplitude produced by loud speech and uniform volume and good quality will obtain at all times. The new transmitter meets this test with a substantial margin.

Carbon leakage is prevented in the new instrument without impairment of transmission by the resilient silk closure for the carbon container previously mentioned.

Aging.—Transmitter design has advanced to a stage where heating at the points of contact in the carbon element need no longer be an important source of aging. Therefore, such aging of the granular material as occurs in a well-designed instrument is limited almost entirely to that resulting from changes in the properties of the granules caused by abrasion of their surfaces when the transmitter is subjected to mechanical shocks such as occur when the handset is placed on the mounting. As in the case of the earlier transmitter, the new transmitter is machine filled[3] with the result that the motion of the granules and the resultant surface abrasion is reduced to a minimum.

The changes in resistance due to the residual aging have little adverse effect insofar as volume is concerned. In fact, the constants of most of the circuits in which the transmitter is used are such that an increase in resistance adds to rather than decreases the electrical output because of the greater amount of power supplied to the transmitter from the central office battery.

On the other hand, an increase in resistance, though small, may prove to be important in certain circuits where a critical relationship between transmitter resistance and the performance of associated apparatus exists. Under these conditions variations in transmitter resistance may result in failure of the associated apparatus to perform satisfactorily if certain limiting values of resistance are exceeded. In determining the limits to be placed on these values account must be taken of all the variables in the circuit in which the transmitter is connected. Obviously combinations of variables of this nature

cannot be dealt with on the basis of averages alone but must include some measure of their range, such, for example, as the standard deviation.[4] The available data indicate that average transmitter resistances and standard deviations which lie within the area bounded by the dotted curve, Fig. 5, will have no adverse effect on circuit

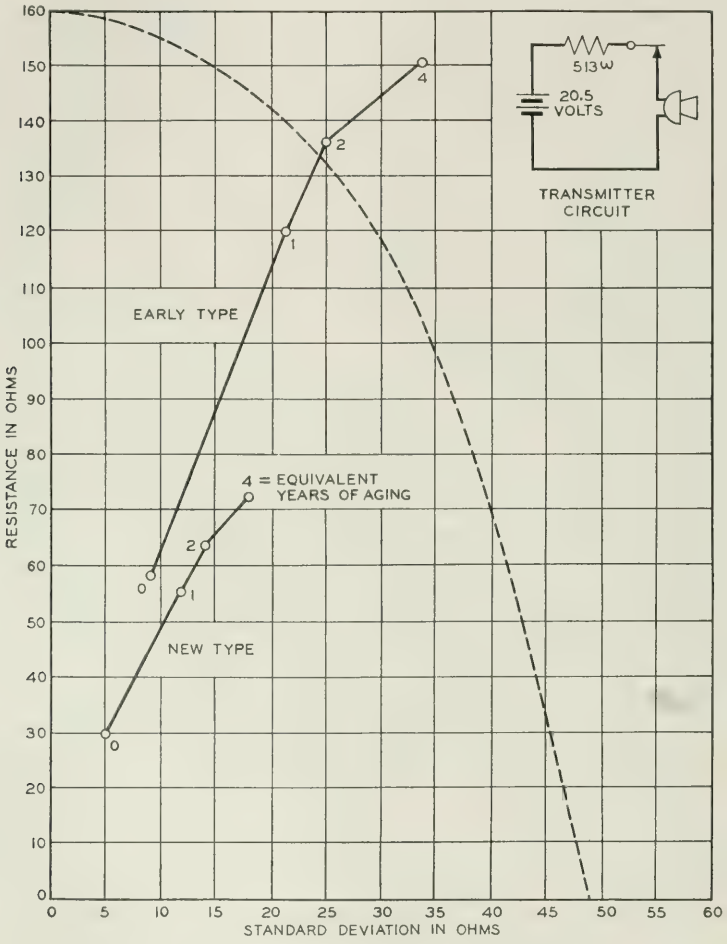


Fig. 5—Limiting values of transmitter resistance.

operation in the Bell System plant. This curve is based on certain marginal circuits of which there are a number in everyday use. An important transmitter resistance in determining the performance of associated apparatus in these circuits is the resistance during the period when the call is being established. This resistance is referred

to as the signaling resistance. As is shown by the solid curves the signaling resistance of the earlier type of transmitter after artificial aging by an amount considered to be the equivalent of four years in service falls outside the acceptable area. On the other hand, the resistance of the new type when aged and measured under identical conditions falls well within the limiting curve and hence not only requires less frequent replacement but also permits greater freedom in circuit design and plant layout.

Moisture condensed from the breath is an important factor in determining the life of a transmitter. A protective membrane is provided in the new transmitter unit which not only is highly moisture resistant but also results in no appreciable transmission impairment. The characteristics of the material employed in this membrane are such that it is not affected by the aging conditions encountered in service such, for example, as the alkaline reaction of water after it has been in contact with phenol plastic parts or tobacco ashes. The exposed metal parts are finished to resist the corrosive action of these agents.

Response.—Reducing the transmitter to an equivalent electrical circuit provides a useful means for analyzing its performance and determining the extent to which the individual parts contribute to its response. Such a circuit for the new unit is shown on Fig. 6.

While the diaphragm can be represented as a lumped mass for frequencies in the region below 3500 cycles per second, it is necessary to consider it as being composed of three masses coupled by stiffnesses in order to represent adequately its performance at higher frequencies. These masses consist of the central portion m_5, the ribbed intermediate portion m_2 and the outer portion m_4. The central portion includes the mass of the movable electrode and is coupled to the ribbed portion by the stiffness s_6 which in turn is coupled to the outer portion by the stiffness s_2. The paper books which support the edge of the diaphragm have a stiffness s_4 and a resistance r_4. Their mass is included in the mass of the outer portion of the diaphragm m_4. The internal resistances of the portions which form the coupling stiffnesses s_2 and s_6 are represented by r_2 and r_6 respectively. A hole is provided in the diaphragm to permit rapid equalization of low frequency pressures of high intensity and prevent damage to the diaphragm and other parts. The mass and resistance of this hole, $m_3 r_3$, are so chosen that their effect on response is confined to frequencies below 300 cycles per second where the station circuit itself is relatively inefficient. The controlling stiffness, s_3, is that of the cavity between the diaphragm and the die-cast frame. As is to be expected the impedance of the carbon granules is a function of amplitude and frequency. However, for the purpose

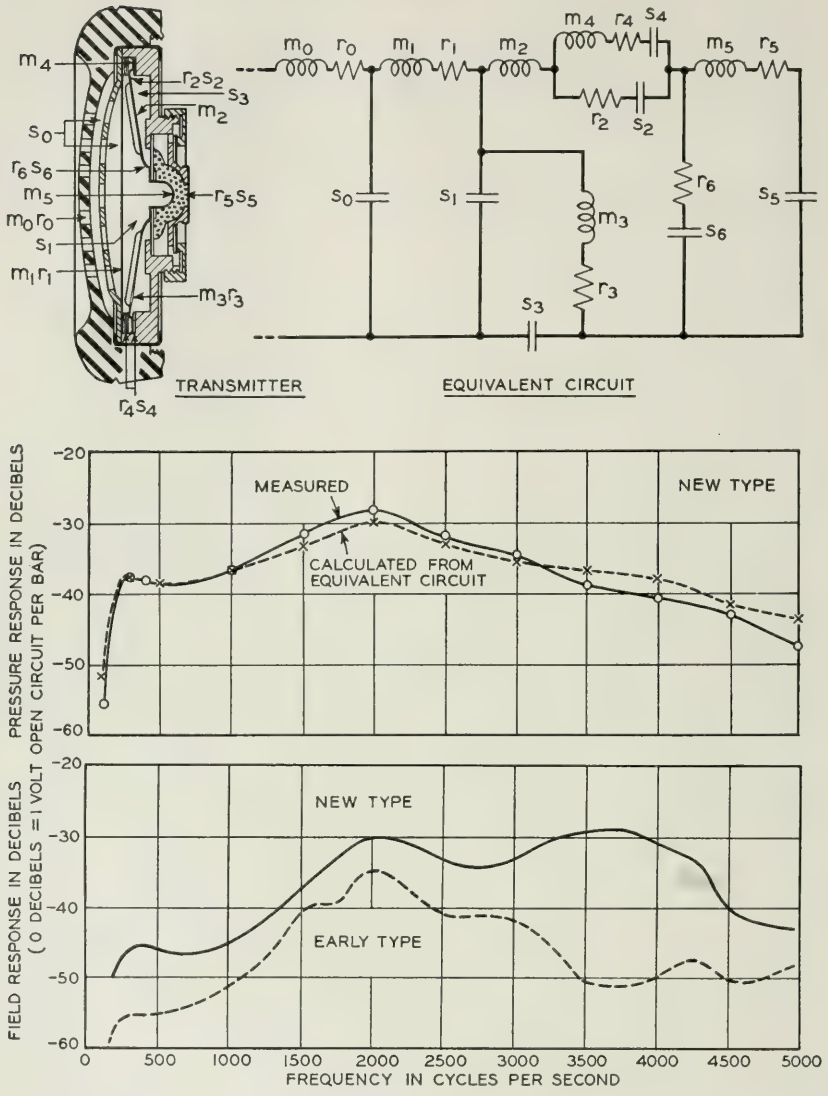


Fig. 6—Pressure and field response characteristics of the transmitter.

of this type of analysis, their impedance characteristics can be represented to a first approximation by a constant stiffness and resistance, $r_5 s_5$. The mass of the carbon is lumped with that of the central portion of the diaphragm. The grid of the transmitter unit proper is provided for mechanical protection only and has holes large enough to have no reaction on response. When assembled in a handset or desk

stand a second grid of insulating material is added. The holes of this grid have a mass and resistance, $m_0 r_0$, which must be taken into account in arriving at an overall picture of the factors affecting response. These holes are coupled to the moisture-resistant membrane, $m_1 r_1$, by means of the stiffness s_0 of the enclosed cavity. The cavity stiffness, s_1, couples the membrane to the diaphragm.

There are two types of response frequency measurements in general use; namely, pressure response measurements in which a constant sound pressure is maintained at the face of the transmitter throughout the frequency range covered by the test, and field response measurements in which a sound field of constant intensity is established at each frequency before inserting the test transmitter. Pressure response is used principally for purposes of analysis whereas field response usually affords a better measure of the performance of the transmitter under the conditions of actual use.

The pressure response of the new transmitter measured with a constant sound pressure at the grid, and the response computed from the equivalent circuit are shown on Fig. 6. The transmitter used in this test was artificially aged by an amount equivalent to two years of service in order to simulate more nearly plant conditions. While the computed curve departs slightly from the measured curve at certain frequencies, due to the inadequacy of some of the basic assumptions, such as those which were made relative to the impedance of the granular carbon, the agreement in general is so good as to provide a powerful tool for predetermining the response characteristics of transmitters under development and a useful method for evaluating the reaction of one element of the transmitter on another. Reducing the transmitter to an equivalent electrical circuit also has proved invaluable in determining the causes of variations in transmitter performance observed during manufacture.

As previously mentioned, the response characteristics of a transmitter under conditions of actual use may differ from those obtained with a constant pressure at its face. In general this is due to the fact that the diffraction effect of the transmitter as an obstacle in the sound field has not been taken into account. While not as readily predetermined as the pressure response, the field response can be easily measured by inserting the transmitter in a sound field of constant intensity. An artificial voice [5] is used in this measurement as a sound source and the electrical input is adjusted to maintain a constant sound pressure at the guard ring at each frequency before inserting the instrument under test. Response curves for the new and early types of transmitter obtained in this way also are shown in Fig. 6.

Both show rising characteristics in the region of 2000 cycles per second. While it is feasible to design a transmitter having a substantially flat field characteristic, experience has shown that the transmission obtained with a transmitter having a rising response in this frequency region more nearly approaches direct speech,[2] when used with a representative line and the new receiver, than that obtained with one having a more uniform field response.

Non-Linear Distortion.—A substantial reduction in non-linearity has been effected in the new transmitter unit.

It is a well-known fact that the slope of the input-output curve of most transmitters is not unity even for sound intensities other than

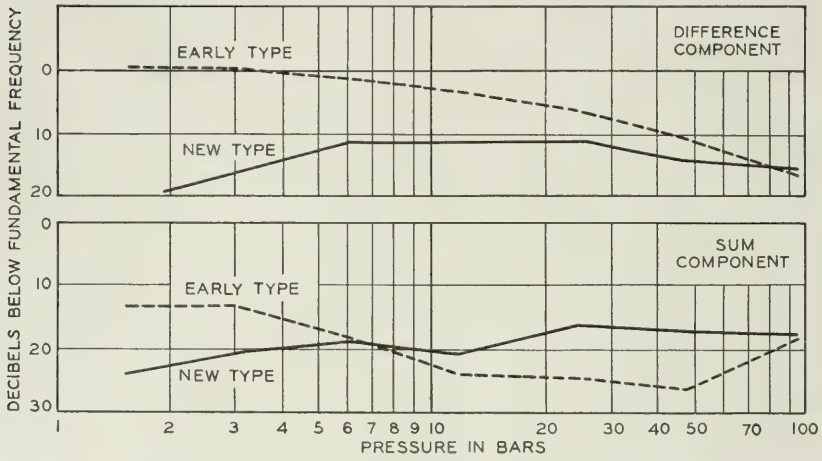


Fig. 7—Sum and difference components as a function of the intensity of the fundamental frequencies.

those at which overloading occurs. However, this departure from non-linearity does not entirely account for the modulation products developed when two frequencies are impressed on the transmitter. Measurements of the sum and difference components in the output of the new transmitter and the earlier transmitter for frequencies of 1500 and 1700 cycles per second are shown on Fig. 7. It will be noted that whereas the difference component produced by the earlier instrument is equal to the fundamental within the speech range, the sum and difference components are both 10 db or more below the fundamental at all the intensities measured in the case of the new unit.

A tendency for the difference component to be considerably more pronounced than the sum component in the case of the earlier transmitter is characteristic of all of the commercial instruments measured

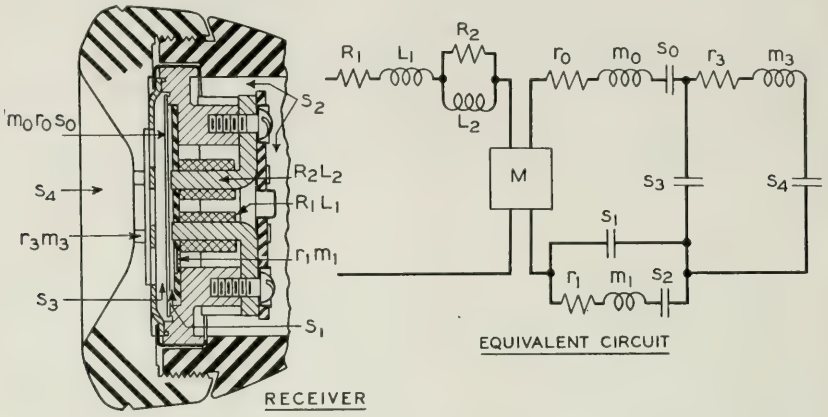
during the development of the new unit. As previously mentioned, this cannot be fully explained by a non-linear relationship between input and output. The available data indicate that it is due to the manner in which the resistance changes cyclically when pressure waves of two frequencies are impressed simultaneously on the transmitter. Similar effects also have been observed in the microphonic action of carbon contacts themselves. Hence it is not unlikely that this is a fundamental characteristic of the carbon itself. If this proves to be true the extent to which an improvement can be effected in the performance of the transmitter will depend upon the ultimate control which can be exerted over the basic properties of the granular material.

Receiver Unit

The new receiver unit is of the bipolar permanent magnet type. The magnetic circuit consists of pole-pieces of 45 per cent permalloy, two straight bar magnets of remalloy and a permendur diaphragm.[6] The magnets are welded to the pole-pieces to form a unit which is mounted on projecting lugs on the die-cast frame. The coils are wound with enamel insulated wire interleaved with cellulose acetate. The pole tips project through a phenol fibre plate which is fastened at the edge to the frame to form a cavity in back of the diaphragm. This cavity is connected to the recess in the handset handle or receiver shell by a hole in the plate. A disc of specially prepared silk covers this hole and provides the required amount of acoustical resistance. The silk fabric is so woven that it does not change in resistance with wetting and drying. The front of the unit is protected by a perforated metal grid which is assembled to the frame by means of a thin ferrule. A disc of impregnated silk is mounted between the grid and the frame to form a screen which prevents the transfer of foreign material from the front to the back of the diaphragm when the receiver is dropped. The grid and ferrule are anodically finished to resist corrosion. The spring contact surfaces are silver plated.

Prior to the introduction of this unit the receivers in general use for telephone purposes in this country and abroad employed simple resonant diaphragms and as a result had response characteristics which were characterized by prominent resonance peaks. As a rule the peak due to the first overtone, as well as that due to the fundamental resonance, fell within the important frequency range. This resonance not only introduced frequency distortion but increased the intensity with which circuit disturbances such as clicks were reproduced. Furthermore, the diaphragms of these receivers were rigidly clamped between surfaces which differed in temperature coefficients of expansion

and heat capacities from those of the diaphragm with the result that the performance of the receiver was erratic and at a given time was dependent upon the temperature changes to which it had been subjected.



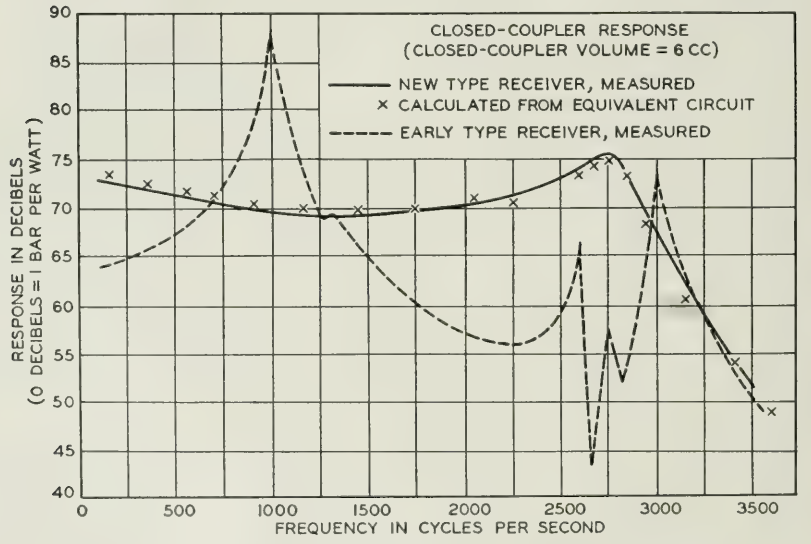


Fig. 8—Closed coupler response characteristics of the receiver.

The new receiver is so designed that:

1. All prominent resonances within the important frequency range have been eliminated and the response within this range materially improved.
2. The effect of changes in temperature has been eliminated.

3. These improvements have been accomplished without sacrificing simplicity of design or introducing features which complicate manufacture of the receiver or increase the maintenance required.

Response.—An equivalent electrical circuit for the receiver and a typical closed coupler response curve are shown on Fig. 8. Referring to this figure it will be noted that there are two meshes in the circuit which contain mass, stiffness and resistance and which control the motion of the diaphragm. One of these meshes consists of the acoustical resistance, m_1r_1, coupled to the diaphragm, $m_0s_0r_0$, by the stiffness, s_1, of the cavity between the diaphragm and the plate which surrounds the pole tips. Included in this mesh is the stiffness, s_2, of the cavity in the handset handle or receiver shell. The other mesh is composed of a cap grid, m_3r_3, and the load, s_4, coupled to the diaphragm by means of the cavity stiffness, s_3. The grid of the receiver unit proper is provided for mechanical protection only and has holes large enough to have no reaction on response. The mass of the resilient screen is small and is lumped with the diaphragm mass, m_0. The electrical portion of the circuit consisting of the winding, R_1L_1, and the equivalent eddy current circuit, R_2L_2, is coupled to the mechanical and acoustical portion by means of the force factor M^7.

The response computed from the equivalent circuit for a number of frequencies is included on Fig. 8. The agreement between this curve and the measured curve is excellent and makes it possible to predetermine the response of the receiver with a high degree of accuracy, and to evaluate the effect on the overall response of the receiver of changes in the constants of the component parts. This type of analysis also has been invaluable as an aid in establishing the causes of variations in response which have been observed during the development and production of the receiver. A measured response curve of a receiver of the earlier type has been added to Fig. 8 for convenience of reference. The improvement in uniformity and range of response is obvious. It will be noted that large gains have been effected for frequencies in the range from 1500 to 3000 cycles per second.

The response of the receiver to a square topped wave affords an excellent measure of frequency distortion. Oscillographic records of the output of typical receivers of the new and earlier types are shown on Fig. 9 for a frequency of approximately 50 cycles per second. The distorting effect of diaphragm resonance is so obvious as to require no comment beyond pointing out that for accurate reproduction of square waves uniform response for an infinite frequency range is required and that the slight rounding of the corners of the wave as reproduced by the

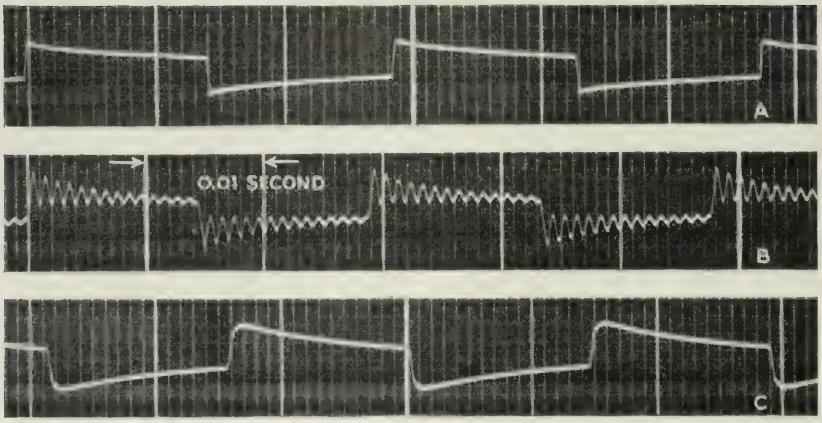


Fig. 9—Response of the receiver to square topped waves.
A—Measuring circuit—no receiver.
B—Early type receiver.
C—New type receiver.

new receiver is due primarily to the falling off of its response above 3000 cycles per second.

The substantially uniform response of the new receiver also renders clicks and other surges much less objectionable. This is due to the fact that the ear does not respond to the peak value of an oscillatory transient alone but integrates the oscillation over an interval at the beginning of the surge, hence the higher the damping the less objectionable the click.

The non-linear distortion produced by a receiver of the new type is negligible in its reaction on transmission, the harmonics in the output being 35 db or more below the fundamental.

Magnetic Circuit.—Inasmuch as the magnetic properties of the diaphragm, as well as its mechanical properties, must be considered in arriving at the preferred dimensions, it was necessary in designing the new receiver to develop criteria which could be applied in determining the optimum relationships between these factors. This study led to the use of the ratio of the force factor to the effective mass of the diaphragm for this purpose. For given magnetic materials in the pole-pieces and the diaphragm and a given air-gap length, there is a pole face area and diaphragm thickness for which this ratio is a maximum. Typical data illustrating this relationship are shown on Fig. 10. The available magnetic materials were studied using this technique and a decision was reached to use permendur in the diaphragm and 45 per cent permalloy in the pole-pieces.

There is a value of polarizing flux for which the force factor of the given magnetic circuit is a maximum. The rate at which the force factor falls off above and below this optimum value of flux is a function of the magnetic characteristics of the materials employed, the length

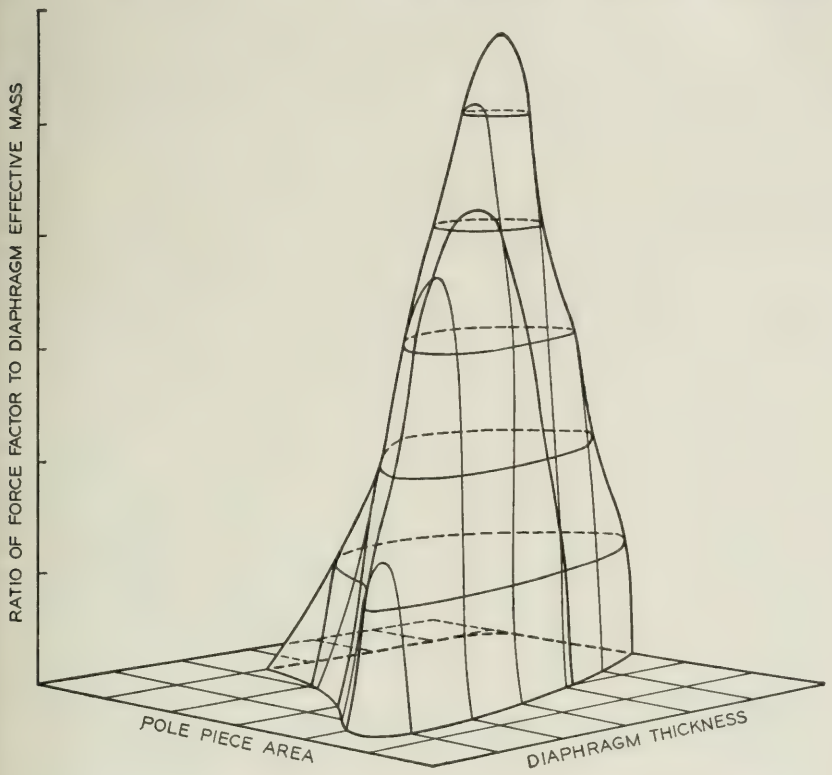


Fig. 10—Force-to-mass ratio as a function of diaphragm thickness and pole-piece area.

of the air-gaps, etc. Without exception the more efficient magnetic circuits have been found to be the most critical as regards polarizing flux. Hence, if wide variations in the efficiency of the product receivers are to be avoided and serious losses due to subsequent demagnetization in service prevented, means must be provided not only for bringing the flux in each receiver to the optimum value, but also for insuring that it remain at this value during the life of the instrument. In order to accomplish this result the magnets of the new receiver are so designed as to overpolarize the magnetic circuit when they are fully magnetized. Equipment is provided for demagnetizing each receiver to its optimum flux value during the assembly process. Receivers

which are not sufficiently overpolarized before demagnetization to resist further demagnetization under service conditions are rejected.

Temperature Effects.—The diaphragm of the new receiver is held in place by the force developed by the polarizing flux and hence it is free to expand and contract independently of its seating surface. This feature renders the performance of the receiver independent of the changes in temperature to which it has been subjected. The force due to the polarizing flux is sufficiently high to prevent rattling at input intensities many times those of loud speech.

Coupling

Although station circuits can be designed which under ideal conditions result in no sidetone, this objective is never fully realized under actual plant conditions, with the result that a part of the electrical output from the transmitter always reaches the local receiver. Whether the electrical coupling between the transmitter and receiver as evidenced by the residual sidetone is of importance from the standpoint of sustained oscillation or "howling," depends upon the degree of mechanical and acoustical coupling between the instruments. Handset and instrument design has advanced to a stage where mechanical coupling need no longer be a problem. On the other hand, as the response of the instruments is improved, the acoustical coupling may become an important item in determining the howling margin. This margin is so large under the conditions where the new handset is being used for transmission purposes that there is no tendency for oscillation or distortion to occur. However, if the handset is placed face downward on a desk or table, an air column is created which resonates in the region of 2500 cycles per second. Inasmuch as this is the region where a substantial improvement in the response of the receiver has been effected, a marked reduction in howling margin results. While there is still sufficient margin to meet all of the requirements of field use, this situation serves to emphasize the fact that such factors as acoustic coupling may limit the transmission improvements which can be effected under a given set of operating conditions.

Effective Transmission

The extent to which the better performance of the new instruments is effective in improving the grade of transmission afforded the telephone user is a complex matter and one which is influenced by such factors as the characteristics of the circuits with which the instruments are associated at a given time, the amount of noise present at the transmitting and receiving stations, the reaction of sidetone on the

loudness with which the user speaks, the distance between his lips and the face of the transmitter, the tightness with which he holds the receiver to his ear, etc. Many of these factors are beyond the control of the engineer responsible for the design of the transmitter and receiver and hence can be evaluated, insofar as their reaction on transmission is concerned, only by tests made under the conditions of actual use.

A method has been devised which makes it possible to rate the overall effect of these factors on transmission in a way representative of the results obtained by the subscribers in their normal use of the instruments.[8] Numerous tests employing this method of rating were made during the development of the new transmitter and receiver to make certain that the course followed in their development would insure the best possible performance under service conditions. Similar tests were also made of the designs selected for production. These tests show that in many respects the new instruments represent outstanding advances in transmission instrument design and performance.

References

1. "Scientific Research Applied to the Telephone Transmitter and Receiver," E. H. Colpitts, *Bell System Technical Journal*, volume 16, July 1937, pages 251–274.
2. "Transmission Features of the New Telephone Sets," A. H. Inglis, this issue of the *Bell System Technical Journal*.
3. "Development of a Handset for Telephone Stations," W. C. Jones and A. H. Inglis, *Bell System Technical Journal*, volume 11, April 1932, pages 245–263.
4. "An Introduction to the Theory of Statistics," G. U. Yule, J. B. Lippincott Company, 1916, page 134.
5. "A Voice and Ear for Telephone Measurements," A. H. Inglis, C. H. G. Gray and R. T. Jenkins, *Bell System Technical Journal*, volume 11, April 1932, pages 293–317.
6. "Magnetic Alloys of Iron, Nickel and Cobalt in Communication Circuits," G. W. Elmen, *Electrical Engineering*, volume 54, December 1935, pages 1292–1299.
7. "Electrical Vibration Instruments," A. E. Kennelly, The Macmillan Company, 1923, page 88.
8. "Rating the Transmission Performance of Telephone Circuits," W. H. Martin, *Bell System Technical Journal*, volume 10, January 1931, pages 116–131.

Transmission Features of the New Telephone Sets *

By A. H. INGLIS

The new telephone instruments now being introduced by the Bell System result in an outstanding improvement in transmission performance in service. The evidence for this, as obtained by comprehensive laboratory and field tests, is presented here together with a discussion of the factors responsible for this superior performance and of the consideration involved in its appraisal.

NEW telephone instruments are being applied in the plant of the Bell System to the deskstand, wallset and handset, and result in markedly improved transmission performance. The new instruments are associated with the anti-sidetone feature which is also applied to the older sets already in plant. The selection of these particular designs from the wide choice made possible by new design technique, materials and manufacturing methods, has been based on developments in the methods for quantitatively rating the relative merits of different designs. In general there has been consistent effort over a period of years to base these ratings primarily on performance in service rather than on laboratory tests.

The factors influencing service performance are so many, and so complicated in their relationship, and are in so many cases difficult or even impossible for the designer to evaluate or control, that their net effect on performance cannot be predicted with certainty by laboratory methods. Of necessity such methods involve a limited selection of primary test conditions, and an even more limited selection from the possible combination of these conditions. This is particularly true in the rating of the transmission performance of a telephone set. Laboratory tests are essential in the study and analysis of design problems, and are invaluable similarly in interpolating, supplementing, and explaining service performance results. In determining the reaction on the user of the transmission features of possible designs, however, the field performance test has been found of first importance in deciding what particular characteristics to include in the new telephone instruments and circuits.

* Presented at A. I. E. E. Summer Convention, Washington, D. C., June 21, 1938

Important Transmission Characteristics of the New Telephone Sets

The specific transmission design features of the new instruments are described elsewhere.[1] The purpose here, therefore, is to discuss primarily the outstanding improvements in performance resulting from the application of the new instruments and the anti-sidetone feature which has been available for some time.

These improvements are

1. Those due to the station circuit, which, as compared with the previous station circuit,
   a—largely reduce the efficiency of the sidetone path between transmitter and receiver without materially affecting the electrical efficiency of the set in transmitting or receiving. This means that sounds, either noise or speech, which are picked up by the transmitter are reproduced in the receiver of the same set at a much lower level.
   b—reduce the susceptiveness for certain types of party line sets to interference with reception by noise set up by power transmission systems.
2. Those due to the physical characteristics of the transmitter and receiver.

Fig. 1—The new handset and deskstand telephone instruments.

Several of these features have been available for some time and have, of course, been introduced into the plant as they became available. The new transmitter and the anti-sidetone circuit, for example, have been standard for some years and have already been installed in large numbers.

Figure 1 shows both the new handset and the deskstand forms of mounting, including all these features as integral parts of their design. The new desk type transmitter and receiver can, of course, be used with wall sets.

The schematic drawing of Fig. 2 indicates the general arrangement of parts in the new station transmission circuit for either type of set.

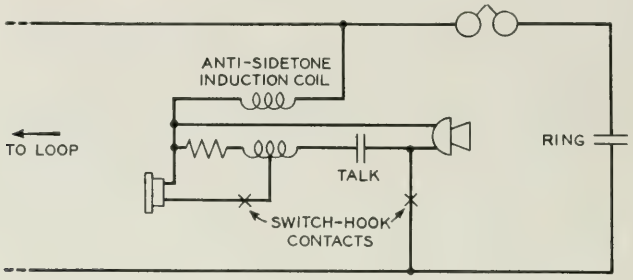


Fig. 2—Schematic transmission circuit of anti-sidetone coil.

In describing the results produced by these transmission features, and the methods employed in measuring and rating these results, it seems desirable to include some discussion of the characteristics of a telephone conversation as distinguished from a direct, face-to-face conversation, so that the various effects of the new circuits and instruments may be seen in as correct relative proportion and as generally comprehensible form as possible.

Some Elements of the Station Transmission Problem

In either a telephone or a direct conversation, successful cummunication depends on the characteristics of the talker and of the listener, and their reactions to each other and to the character of their surroundings. In a direct conversation such, for example, as across a desk, the environment is in general the same for both talker and listener, and their ears are materially aided by their eyes. In a telephone conversation, however, not only may the surroundings of talker and listener be entirely different, but a third element, the telephone system, is added to the environment of each user, which complicates his reaction, not only to his own surroundings, but also to the other party to the conversation. Furthermore, for obvious economic reasons, the natural binaural reception of direct conversation, with its advantages in discriminating between sounds from different directions, is replaced in the telephone conversation by a monaural medium.

Fundamental differences of this kind between telephone and direct conversation must be taken into account in the design of a telephone transmission system if satisfactory results are to be obtained. For example, the talker is accustomed in a direct conversation to regulate his talking volume by what he himself hears under prevailing noise conditions (which incidentally are the same for the listener), by the ease with which he hears the other party, and by the ease with which the listener appears to hear him. By experience, under ordinary conditions, the first factor mentioned, the loudness with which the talker hears himself, probably comes to be the primary control on his talking volume.

These various factors also serve to regulate talking volumes in conversation by telephone, but their magnitudes and the relations between them differ from the condition of face-to-face air path conversation and vary from one type of telephone connection to another. For example, the "sidetone" of the telephone set, being materially higher than the air-path sidetone, deceives the talker, not only by making him think he is talking louder than he really is, but also by apparently modifying the noise conditions under which he is talking in the pickup and amplification of room noise by his telephone transmitter. Since, in addition the efficiency of the telephone circuit itself may be different in the two directions of transmission, the loudness heard by one party may differ more from that heard by the other than in the case of air transmission. Then, too, noise conditions may be and frequently are quite different at the two ends of the telephone circuit. Figure 3*A* shows the probability of noise of various average intensities at subscribers' stations as determined by several surveys covering a large number of locations. On the assumption that any one of the stations represented by these data may with equal probability call any other one, Fig. 3*B* has been computed, showing the probability of noise at the two stations of a telephone connection differing by more than a certain amount. It will be noted that there is about an even chance of the noise at the two ends differing by more than 12 db. In view of these differences, a person's judgment of how well he is heard and understood can not be as direct as in the case of air transmission.

In addition, the transmission over the commercial telephone system affects the quality of the received speech more than the usual room surroundings in air-path transmission. While acoustic resonance and reverberation in a room do distort speech, in the extreme case to a point where understanding may be difficult, such a condition is distinctly unusual. Equal freedom from distortion in a telephone system

is a more difficult and expensive condition to obtain than in direct conversation a few feet from a listener. Something less than perfect reproduction must suffice, for the present anyway, if costs are not to be prohibitive.

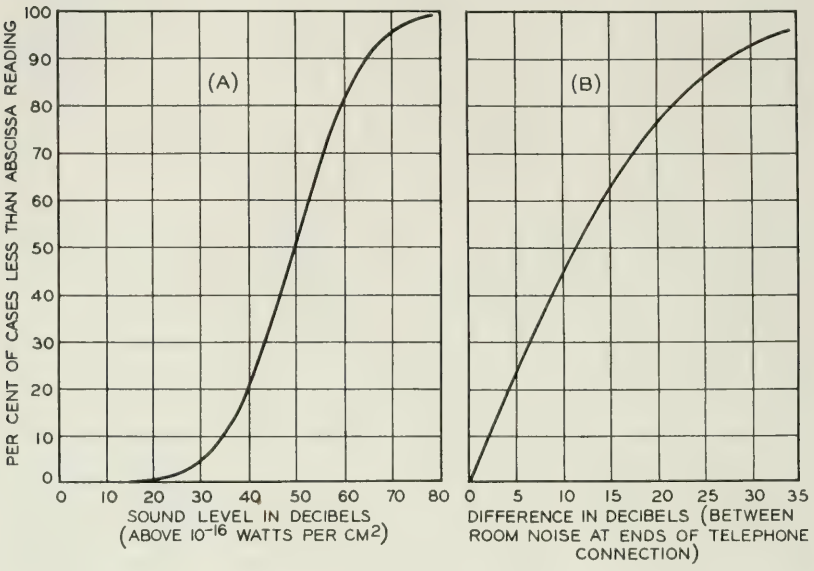


Fig. 3—Noise conditions at telephone stations.

All of these differences involve the acquiring by the user of a set of telephone habits which differ from those he has acquired in direct conversation. The problem of the transmission design of a practical telephone system requires, then, for a satisfactory solution, not only a determination of the proper speech levels to be delivered, and of the sidetone characteristics which will, under the conditions of a telephone conversation, give optimum results with the noise encountered, but also a decision as to what particular frequency range and characteristic to choose. Properly designed, a telephone transmission system should minimize, to the degree consistent with costs, its inherent differences from direct conversation, and make it easy for the ordinary user to get, without undue effort, results which are satisfactory to him in comparison with direct conversation.

In the earlier days of telephony, the problem presented appeared much simpler. It was, in effect, uni-dimensional, calling primarily for more efficient instruments and circuits; more and more power delivered to the listener's ear. While methods for the control of sidetone were

not unknown, the importance of such control was not fully appreciated. Little choice was available in the quality of reproduction provided by transmitter and receiver, because of meager design knowledge.

Relatively recent and quite rapid developments in knowledge of the problems involved, in materials, methods, and measuring facilities, have now presented the necessity for a solution in essentially a three-dimensional form. These three dimensions may be described as volume, noise and quality. The solution of the problem on this basis is obviously more difficult, and has required the development of methods for quantitatively evaluating and rating their characteristics in terms of some common yardstick.

Methods of Rating Transmission

For reasons already suggested, such a yardstick must be based on service performance—on the results obtained by actual users in the course of day-to-day telephone service.

Extensive investigation has indicated that the best comparative measure of this transmission performance in local exchange service is to be found in the time rate of the occurrence of repetitions required by subscribers for understanding telephone conversations.[2] Or, more explicitly, when two transmission conditions have the same repetition rates, all other service factors being equal, these conditions are taken to be equal with respect to transmission performance. Where two conditions are not alike it is usually possible to evaluate the difference in the repetition rates for the same users by inserting distortionless loss in the better condition until both have equal repetition rates.

Thus, by taking as a reference a typical telephone circuit of specified make-up, the effects of various factors such as distortion, noise, attenuation, sidetone, or type of instrument, may all be expressed in the common terms of the reference circuit trunk which will give the same repetition rate.

Instead of making this adjustment in every case for the purpose of evaluating the relative performance of different test conditions for the same users, the evaluation may be made rather closely over a limited range by the following typical relation derived from repetition observations on circuits containing trunks, the losses of which were varied over a range of values.

$$\text{db} = 50 \log_{10} R_1/R_2,^3$$

where R_1, and R_2 are the repetition rates of two conditions under comparison, and the db figure is the change in the reference trunk which has the same relative effect on the repetition rate.

Such a method is, of course, somewhat cumbersome, and requires a large amount of data to iron out random variations and individual peculiarities of little general interest. But as the fundamental rating method, supplemented by laboratory test, it has been systematically used in studying the value of the anti-sidetone circuit and in selecting instrument characteristics.

Supplementing the repetition observations, it has been found useful in service rating to obtain data on speech levels delivered to the line for each condition observed. This has been done with the volume indicator, a vacuum tube voltmeter so designed that the reading is approximately proportional to mean syllabic voltage.[4] The information thus obtained is useful not only in analyzing the results of service tests but also in determining typical values for speech levels, necessary for laboratory tests.

Laboratory tests are of two general types: objective measurements, and subjective tests. Transmission measurements cover a wide field with objectives ranging from the physical analysis and study of different designs, to the determination of overall performance characteristics of structures and systems. It is these latter tests that we are more particularly interested in here, as most descriptive of the physical properties of importance in providing telephone transmission service.

Subjective tests in the laboratory may be said to be midway between physical measurements and field performance tests. Made under controlled and somewhat artificial conditions, they indicate quantitatively the capabilities of a telephone system in transmitting articulate speech under the particular conditions of the test. They cannot, of course, indicate the relative probability of occurrence, and hence importance, of these different conditions, nor predetermine how well the subscriber will avail himself of the capabilities provided.

Consideration of some of the results of investigations in both laboratory and field will do much to explain the rather large transmission improvement realized by the introduction of the new sets in actual service, particularly if examined with the conditions of a direct conversation as a basis of comparison.

The Station Circuit

There are two characteristics of the new station circuit of particular importance from a transmission standpoint.

Reduction of Sidetone

The first is the anti-sidetone induction coil through which the transmitter and receiver are coupled to the line. This coil comprises,

in addition to three transformer windings, a balancing network. The circuit, made up of the four elements: transmitter, receiver, line, and network, coupled by the transformer, functions in such a manner that the transmitter and receiver are in conjugate relationship, i.e., voltages produced by the transmitter are balanced out and do not affect the receiver. Theoretically, such a circuit, with pure resistance elements, can be perfectly balanced at all frequencies with complete elimination of sidetone, and at the same time be as efficient as can any transformer coupling in an invariable telephone set,[5] for the transfer of power from the transmitter to the line, and from the line to the receiver.

This type of circuit is not new in principle, and many varieties are known and have been described.[6] Many of these arrangements, for one reason or another, are not suitable for application. Some, for example, call for impedances of transmitters or receivers differing widely from those available. Certain others are not economical for common battery service, where the transmitter must receive its battery supply from the line. Still others require relatively complicated and expensive cording and switchhook arrangements. The circuit which has been chosen for general common battery subscriber station application, and shown schematically in Fig. 2, is not only as simple and as easily adapted to Bell System conditions as any, but permits a coil design which is economical to manufacture as well as efficient in performance. Other types of anti-sidetone circuit have been adopted for local battery station service and for operators' telephone sets.

The theory of operation of this anti-sidetone circuit has already been discussed elsewhere.[7] It is intended here to show the general purposes of the application, some of the considerations involved in the design, and the kind of results accomplished.

While in theory complete elimination of sidetone is possible, as well as ideal efficiency of transformation, in practice neither objective can be entirely realized. The unavoidably wide variations in line impedance looking from the set, ranging from high positive to high negative phase angle, and from a few hundred to more than a thousand ohms in magnitude, together with other practical departures from ideal conditions, necessitate a choice between a high degree of sidetone balance and the standardization of a minimum number of coil designs. The variations in loop length and resistance, by their effect on transmitter battery supply, and consequently on transmitter resistance, furthermore cause variations in the absolute transmitting and sidetone efficiency of the terminal set, which must be taken into account in the station circuit design.

The actual design chosen is so arranged as to favor sidetone balance

on average and shorter loop conditions where transmitter battery supply is greater, with consequent higher sidetone, and to favor transmitting and receiving efficiency on longer loops where battery supply is low. Since loop losses are greater for transmitting than for receiving because of transmitter battery supply loss, the ratio of the transformer is such as to favor the transmitting efficiency of the set somewhat in comparison to the receiving efficiency. This has the advantage of raising the transmitted speech level further above line noise. The same idea, of course, was followed in the design of the sidetone set.

The resultant anti-sidetone circuit adopted and here discussed, as compared with the sidetone circuit previously in general use, when equipped with the same transmitter and receiver and on the same loop and trunk, reduces sidetone on the average by about 10 db. Under the most unfavorable conditions of use, the reduction is unlikely to be less than about 7 db compared with the corresponding sidetone connection. Under the best conditions of balance encountered the reduction may be as much as 12 db. On the effective basis of transmission the average net improvement in transmission which results is about 6 db.

From the electrical circuit standpoint alone, the efficiency of the anti-sidetone arrangement is below that of the sidetone set in the order of about one or two db in transmitting and in receiving, which is necessitated by the limitations of practical design and circuit conditions discussed above.

Figures 4*a* and 4*b* show for transmitting and receiving, respectively, the difference in efficiency, with respect to frequency, of the anti-sidetone set from the sidetone set, each with the same instruments. Two subscriber loop and trunk conditions are shown: an average loop and trunk, and a long cable connection.

Figure 4*c* shows the variation in sidetone reduction with frequency, of the new set as compared with the corresponding standard sidetone set, for the same two circuit conditions as above. The curves are indicative of the effect of variation of circuit impedance on sidetone balance, in changing not only the magnitude, but the frequency range in which the best balance occurs.

Data of this sort alone do not, of course, indicate the relative transmission performance of the two sets. The beneficial effect on the telephone user of the large reduction in sidetone must be evaluated on the same yardstick as the losses in transmitting and receiving efficiencies which, in the practical case, accompany this reduction in sidetone. McKown and Emling have shown the effect of changes of this sort on the results obtained by the ordinary telephone user, in

terms of net effective transmitting and receiving loss, as determined by service observations.[8] Their data, shown in Fig. 5, are relative to the sidetone of a reference set. The heavy solid lines are the original experimental data, the dotted extensions to these curves being extrapolated.

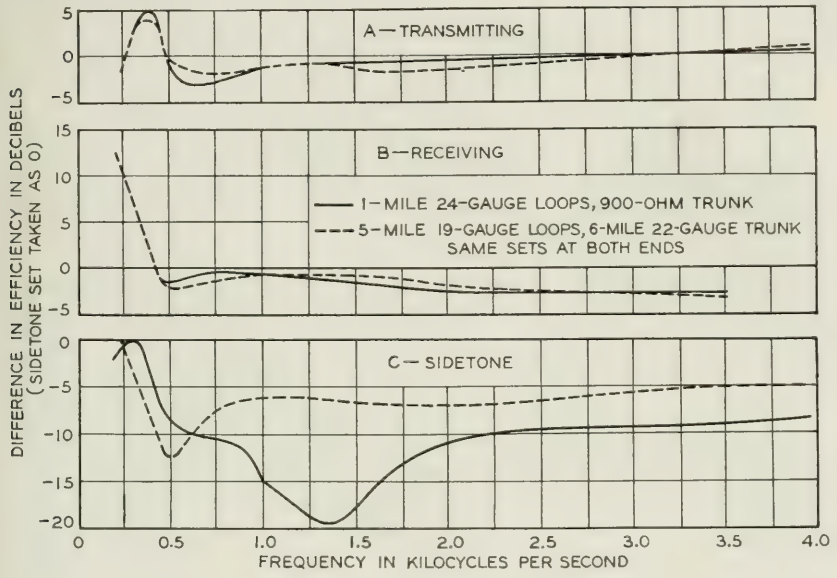

A—TRANSMITTING
B—RECEIVING
1-MILE 24-GAUGE LOOPS, 900-OHM TRUNK
5-MILE 19-GAUGE LOOPS, 6-MILE 22-GAUGE TRUNK
SAME SETS AT BOTH ENDS
C—SIDETONE
DIFFERENCE IN EFFICIENCY IN DECIBELS (SIDETONE SET TAKEN AS 0)
FREQUENCY IN KILOCYCLES PER SECOND

Fig. 4—Circuit efficiency of anti-sidetone circuit vs. sidetone circuit.

In addition to the original ordinates, others are shown which are of interest. They are based on the results of loudness balance tests, and while not perhaps of great precision, do approximately indicate the relationship of the sidetone of a telephone conversation to that for direct speech, and illustrate the differences in the effects of sidetone for transmitting and for receiving.

On each curve are indicated the average sidetone value of the standard sidetone and the new anti-sidetone set, each, as before, with the new transmitter and receiver. There is also shown the range of sidetone for each type of set, within which practically all service conditions will fall. This indicated range takes into account not only variations in sidetone balance due to line impedance variations, but changes (with loop resistance) in battery supply to the transmitter. It should be noted that in only a few cases is the absolute sidetone of the anti-sidetone set on the worst sidetone conditions, as high as or higher than that of the sidetone set on the best sidetone

conditions, and then only by a small amount. Furthermore, in spite of the wider variations in sidetone of the anti-sidetone set, these variations are over a range such that the resultant variations in effective losses are smaller than for the sidetone set.

Considering Fig. 5*a*, it will be noted that for either sidetone or anti-sidetone sets, the sidetone is louder than for natural speech

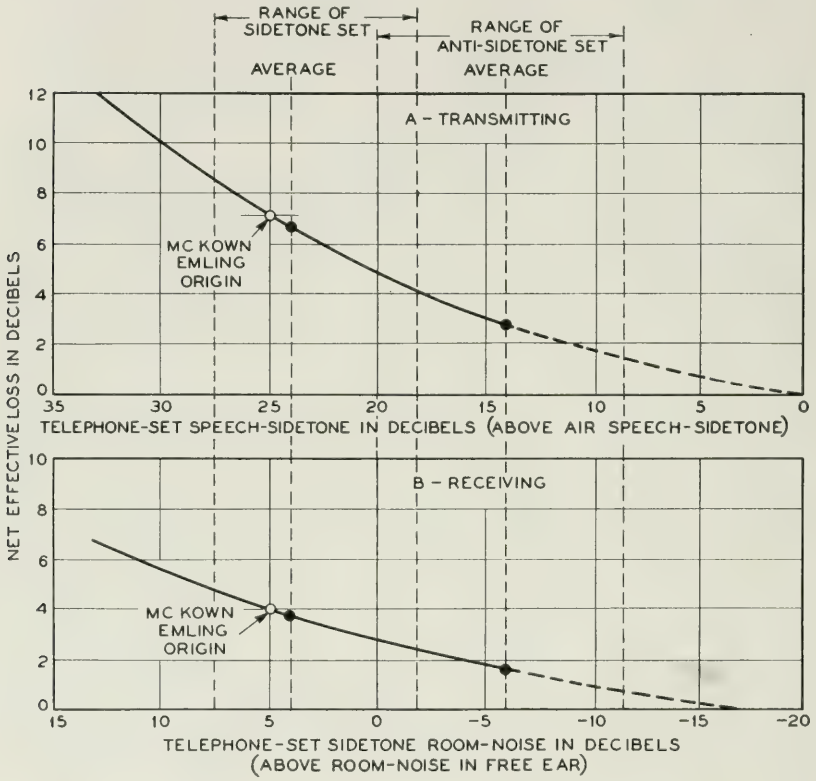

RANGE OF SIDETONE SET
AVERAGE
RANGE OF ANTI-SIDETONE SET
AVERAGE
A - TRANSMITTING
MC KOWN EMLING ORIGIN
NET EFFECTIVE LOSS IN DECIBELS
TELEPHONE-SET SPEECH-SIDETONE IN DECIBELS (ABOVE AIR SPEECH-SIDETONE)
B - RECEIVING
MC KOWN EMLING ORIGIN
TELEPHONE-SET SIDETONE ROOM-NOISE IN DECIBELS (ABOVE ROOM-NOISE IN FREE EAR)

Fig. 5—Effects of sidetone on user of telephone.

sidetone, which, as noted before makes the user think he is talking louder than he actually is. The average sidetone reduction of 10 db for the anti-sidetone set results in less of this restraint on his talking level, with a resultant net effective gain in transmitting of about 4 db compared with the sidetone set.

In receiving, Fig. 5*b*, sidetone introduces an effective loss by the reproduction in the telephone ear of room noise picked up by the transmitter. It will be noted that for the anti-sidetone set, the reproduced noise is in general appreciably lower than the room noise

itself. Inasmuch as room noise also interferes directly with received speech in the telephone ear by leakage under the receiver cap, the contribution to the total noise of the sidetone pickup of the anti-sidetone set is in most cases small. This is not true for the sidetone set, where in many cases the sidetone noise may constitute the principal interfering noise. The resultant net effective gain in receiving is about 2 db compared with the sidetone set.

This is a good illustration of the type of information which can be obtained only from a field study. For example, the relationships indicated on Fig. 5 are dependent on how far away from the mouthpiece of the transmitter and at what level the speaker talks, and on how tightly to his ear he holds the receiver. These in turn are resultants of all the conditions of the particular telephone conversation. If incoming levels are so high as to be uncomfortable, the receiver may well be held farther away from the ear. In that event, of course, the sidetone conditions of the set become relatively less controlling. The weight, size, and shape of the instrument in his hands may similarly affect the subscriber's use of it, the results he gets, and the relative importance of various factors of telephone design.

For such reasons, not only must laboratory performance tests be supplementary and subsidiary to field tests, but additional field tests must be the basis for determining the effect of any major changes in design, whether or not those changes are electrical, acoustical, or purely mechanical.

Considerations of this sort emphasize the importance of having clearly and explicitly in mind the conditions and relationships of direct conversation, as a general reference for the interpretation and explanation of the effects of telephone design on telephone conversations. The sidetone ordinates of Fig. 5, for example, not only suggest the difference in function of the anti-sidetone circuit in transmitting and receiving, but also emphasize the fact that the overall sidetone resulting from the combination of circuit, instruments, and method of use, is the important factor rather than the sidetone circuit efficiency only. Such matters are easily lost sight of, if design is not properly coordinated in its correct perspective.

The reduction of sidetone provided by the anti-sidetone sets is of further advantage in two rather different ways.

In attaching a transmitter (which is an amplifier) and a receiver, to a common handle which mechanically couples the two, a condition is set up in which the gain under certain conditions may exceed the loss in the path made up of handle, air, and electrical sidetone circuit. Sustained oscillation, or howling, will then result between transmitter

and receiver. Even if this point is not reached, but is approached within 6 db or so, impairment of quality results from incipient oscillation. The greater sidetone circuit loss of the anti-sidetone circuit provides an additional margin of safety against any such condition.

The granular carbon of the transmitter, and the design of the transmitter itself must be carefully controlled, or serious noise—transmitter "burning"—will cause noise in the receiver of the set. The mechanical and electrical wear and tear of service tend to make this transmitter noise worse. In the new transmitters this "burning" has been kept at a low inherent value throughout life. The anti-sidetone circuit, however, provides a margin of safety against the small amount remaining, so that with this set there is less likelihood of transmitter noise causing impairment of reception.

Reduced Susceptiveness to Interference

It will be noted from the schematic circuit drawing Fig. 2 that two condensers are used in the new sets, one in the anti-sidetone transmission circuit, and a separate one with the ringer. In some types of party line practice the ringer of the set is connected for some parties from one side of the line, and for the others from the other side to ground.

Figure 6 shows schematically two such ringing arrangements during the conditions of conversation, 6*a* as used in the new sets, and 6*b* with one condenser common to transmission and ringing circuits. It will be noted that in the standard circuit adopted (Fig. 6*a*), if any longitudinal noise voltages exist between the central office and station grounds, there is an equal voltage drop from each side of the line to ground through the ringing paths (assuming the two ringer condenser paths to be identical). The voltage drop across the terminals of the talking set is therefore zero and no noise results.

If the arrangement of Fig. 6*b*, corresponding closely to previous designs, were used, however, this condition would not obtain. The condenser of the station in use being common to the transmission circuit as well as the ringing circuit, the noise voltage drop across this condenser is introduced in the transmission circuit. In addition there are other paths to ground from each side of the line through the transmission circuit which are not of equal impedance. The net result is a residual noise current through the receiver of the talking circuit.

In the actual case, the impedance of all ringers and condensers is not identical and there are often more parties connected to one side of the line than the other. Even under these relatively unfavorable conditions, however, the two-condenser arrangement adopted reduces

the susceptiveness of the set to interfering noise by as much as 15 db. A further material reduction is realized by the high impedance of the ringer used in the new sets, so that in most cases interfering noise at grounded ringer stations will not differ materially from that at individual stations where the ringer is bridged across the line.

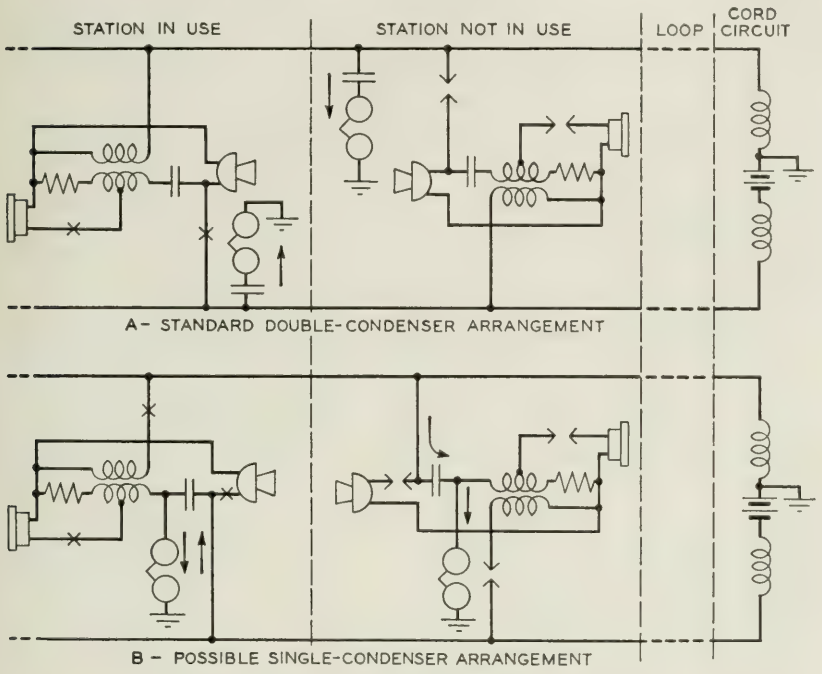


Fig. 6—Ringing arrangements for party line service.

It is interesting to note that this improvement is realized at little additional cost, since the transmission condenser, which must be of relatively high capacitance, is permanently bridged by the transmitter, so that it is protected from exposure to any large voltages, and may be of cheaper construction and smaller in size than would otherwise be the case. The ringing condenser on the other hand, while it must be constructed to withstand higher voltages, may be of relatively small capacitance, which gives more uniform and better ringing and dialing performance.

Characteristics of Transmitter and Receiver

Since the individual design characteristics of the new transmitter and receiver are discussed elsewhere,[1] attention here will be centered on the overall effects of these characteristics in the complete trans-

mission system, as indicated both by laboratory and by field test. As stated before, the problem may be more or less arbitrarily separated into three correlated problems—volume, quality and noise.

As in the case of sidetone, these problems appear, perhaps, more nearly in a proper perspective if considered in comparison with the corresponding factors in a direct conversation. It must be remembered that telephone service does not consist in the provision of a mechanism, per se, but in the provision of facilities for conversation, to which the mechanism should be incidental, however important. Since the inherent conditions of such a conversation are quite different in many respects from those of a direct conversation with which, consciously or unconsciously, it will be compared in its overall results, the parallelism in detail should not be too exact. Departures from the conditions of direct conversation in certain respects which are relatively unavoidable, may be best compensated for by deliberate departure in certain other respects. For example, the physical absence of one party to the telephone conversation, and the monaural nature of such a conversation, may be partially compensated for by delivering to the ear of the listener a somewhat higher speech level than he is accustomed to in direct conversation. The limitation of frequency band width imposed on the telephone medium, largely for economic reasons, may be minimized in its effects if the transmission characteristics in the available band are other than a facsimile of the corresponding band in direct conversation. All such measures must be employed with knowledge of their effect on the ultimate objective, that the telephone conversation may be easy and natural.

General Requirements

It is easily seen that for any particular overall frequency characteristic of a telephone transmission system, there are practically an infinite number of ways in which it can be split up between transmitting and receiving characteristics. From this standpoint alone, then, there is no particular "best" transmitter or receiver frequency response. From other standpoints, however, certain general types of individual characteristics, both in frequency and efficiency, are to be preferred to others, particularly when considered in their practical application to an already existing telephone system. It has been pointed out [9] that in general, development has been toward a telephone system where both transmitter and receiver are relatively uniform in their frequency characteristics. Induced noise appears to be so evenly distributed with frequency that such response would not appear to magnify the interference problem.

Transmitting efficiency should be as high as required to keep the speech well above induced noise but not so high as to cause excessive crosstalk into other telephone circuits. The maximum desirable received level is determined for a given telephone system by the limitations of the human ear in accepting with comfort speech levels above a certain intensity. Finally, the practical necessity of working as satisfactorily as possible in conjunction with the telephone transmitters, receivers, and sets in the existing plant during the period of transition, places a practical limitation on the amount of change that is desirable in relative levels of either transmitting or receiving.

With regard to frequency range, previous work [10] indicated the desirability of designing circuits to transmit frequencies from 200 or 300 cycles up to about 3,000 cycles. Gains in articulation and naturalness are realized by increases in this band width, but are progressively smaller for successive equal increments in frequency. A 3,000-cycle band properly used gives good transmission both in articulation and naturalness, but frequency limitation is essentially an economic one, subject to change as conditions change. Recent work on the new multiple channel carrier systems has indicated justification in these systems for providing a somewhat wider band, from about 150 to about 3,500 cycles.[11]

Overall Frequency Response

In describing the frequency characteristic of a transmission system it has become customary to refer to it as more or less "flat," where "flat" is assumed to be synonymous with "perfect" as far as the relative transmission of various frequencies is concerned. In measurements of the elements of an electrical circuit, from which this terminology came, the word is useful since, when the measurements are properly made, at any rate, the basis of comparison implied by the word "flat" is generally understood. This is also true, although probably to a more limited extent than is generally realized, when the term is applied to electro-acoustic transmission systems, where free progressive, plane air waves of various frequencies are transferred to an electrical system, or vice versa, by means of microphones or loud speakers.

In the case of a telephone system, however, where a transmitter is placed close to the lips, and a receiver directly to one ear, and where the air waves are not free progressive, or plane, use of the word "flat" implies a basis of comparison which is not self-evident. Much effort has been given recently to establishing an appropriate reference system, sufficiently simple in concept and ease of specification, to be useful in this connection. The result of this work has been a reference telephone

system which, when spoken into, would give the listener in all respects essentially the identical sensation he receives in one ear when facing the speaker directly, with an air path one meter long between the speaker's mouth and the listener's ear, in surroundings without reverberation or noise. Such a reference transmission system has tentatively been called an "orthotelephonic" system.

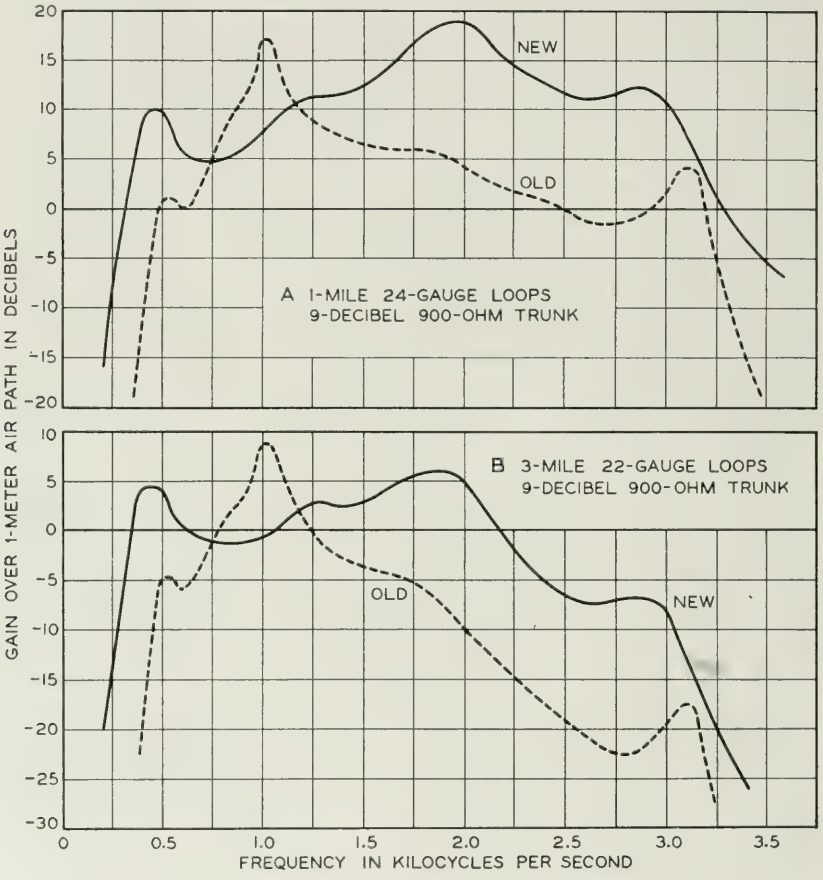


Fig. 7—Overall orthotelephonic frequency response of Typical Telephone Connections.

The point of interest here is that when measured by any suitable objective method, the frequency characteristic of this "orthotelephonic" telephone reference system is not "flat" by a considerable amount in the ordinarily accepted usage of the term. This departure from "flatness" is caused by such factors as the frequency directive

characteristic of the mouth, cavity resonance of the ear, and disturbance of the sound field by the head. The individual contribution of some of these factors is not as yet definitely determined.

Furthermore, for reasons mentioned, it is not self-evident that a practical telephone system of limited frequency range should be "flat" with respect to the corresponding frequency band in this more or less basic orthotelephonic system which is not limited in frequency range. Having decided on the band width that is desirable and justifiable, it must still be determined, therefore, what particular frequency characteristics are preferable in this band.

In selecting from the many possible choices, the particular frequency response that seems best, several factors must be taken into account. This has been done by a study (under the conditions of actual service) of the relative results of several different experimental instrument designs, varying in frequency characteristics. The overall frequency characteristics of the resultant choice are indicated for two typical circuit conditions in Fig. 7. These measurements were made with the artificial mouth and ear [12] and are plotted with reference to corresponding measurements on an orthotelephonic reference telephone system. For comparison, the results of similar tests of the earlier Bell System handset [13] are shown also.

In considering these overall telephone system frequency response characteristics in the light of previous discussion, there are several points of interest:

1. The large increase in response at both higher and lower frequencies with respect to the older handset, which in itself was a notable advance in this respect over previous types. This increase amounts to 10 db or more from about 200 to 500 cycles and from about 1,700 to 3,000 cycles. This wider frequency range gives better naturalness of reproduction.
2. The type of the response. The general uniformity and absence of any marked resonance or irregularity is obvious. For either average or long loops the entire band from about 300 to over 3,000 cycles lies within a range of 15 db. It will be noted, however, that, for the average condition, the response at the higher frequencies (1,500–3,000 cycles) is distinctly above that for the frequencies below 1,500 cycles. This characteristic aids materially in the understanding of the low intensity consonant sounds. The response on the longer loops would undoubtedly be correspondingly better if the high frequencies were raised so that the overall characteristic more nearly resembled that for the average condition shown. It should be remembered,

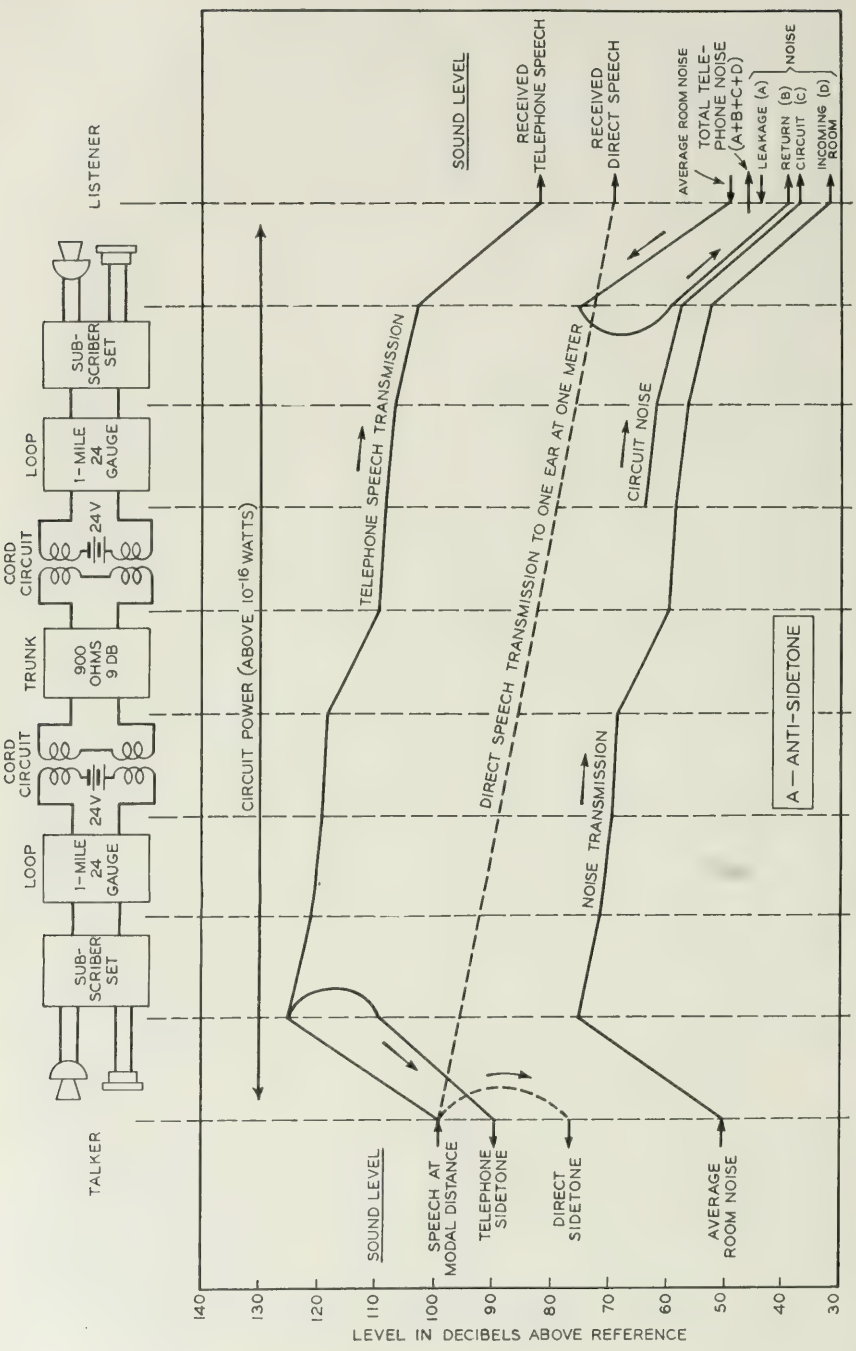


Fig. 8—Power level diagrams of typical telephone connections.

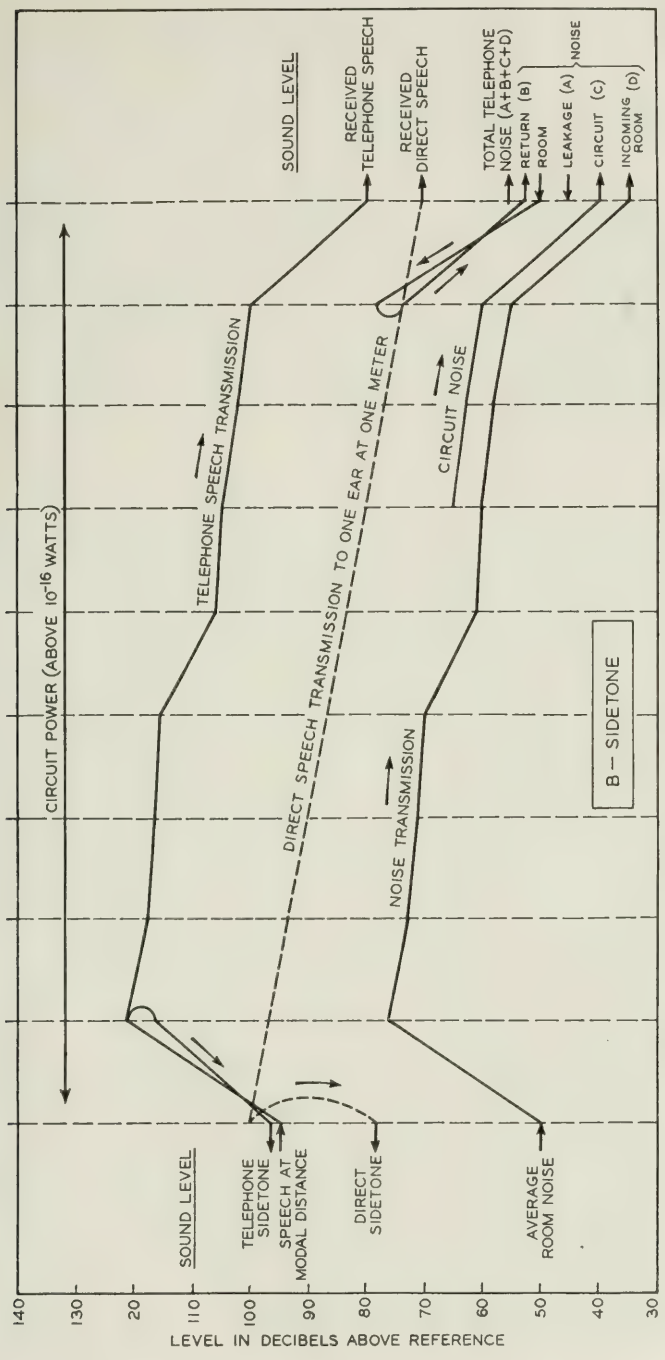


Fig. 8.—*Continued from page 376.*

however, that for a standard set, to be used on all loop conditions, a response designed solely for the long loop, with its large loss at high frequencies, would be distinctly "tinny" and disagreeable in quality on average or short loops.

3. The materially smaller losses in the transmitted band for the average telephone connection than for the orthotelephonic system. Even for the long loop conditions the losses are no greater than for this orthotelephonic system up to about 2,200 cycles. In other words, for a large majority of telephone calls the received speech level will be higher for the same talking level than in the case of direct conversation at one meter distance. The desirability of this in a monaural system of limited frequency range has already been indicated.

Combined Effects of Circuit and Instruments

The part played by the anti-sidetone circuit in permitting better utilization by the subscriber of the capabilities of the telephone system, in increasing the level of received speech, and minimizing the effect of noise is summarized by the illustrative power level diagrams in Fig. 8, using typical values of room and line noise and other conditions.

The data shown were obtained by objective measurements with the artificial mouth, sound level meter, volume indicator, and artificial ear. Frequency weightings appropriate to the levels involved were employed in the sound level measurements.

Figure 8*a* shows, for the anti-sidetone set, the losses between the talker and listener under average room noise and circuit conditions. Figure 8*b* gives corresponding information for the sidetone set. For comparison, an approximate level diagram is shown also for direct speech. The relative speech levels at the transmitter for the two telephone conditions were adjusted to give volume indicator readings on the line in accordance with the results of service observations. The upper curves in each drawing are for speech and the lower for noise picked up by the transmitters.

At the transmitting end the lower sidetone of the anti-sidetone set results in higher talking levels, with about 4 db higher speech level at the input to the circuit. In the overall circuit this gain is increased by the receiving-end effect of the anti-sidetone circuit in minimizing the effect of room noise.

The total noise in the telephone ear, as shown on the drawing, has as its principal contributing factors:

1. Leakage of noise under the receiver cap.

2. Noise picked up by the transmitter and returned via the sidetone path and receiver to the listener's ear—termed "return noise."
3. Circuit noise.
4. Room noise picked up at the far end and transmitted over the circuit.

The different relative contribution of the "return noise" for the two sets is of interest. The net result is a total noise in the telephone ear, lower than the actual room noise for the anti-sidetone set, and higher for the sidetone set.

For the circuit and room noise conditions shown, the ratio between received speech level and noise is about 25 db for the sidetone condition, and 35 db for the anti-sidetone. The corresponding ratio for air transmission to one ear under the conditions shown is in the order of 20 db.

Results of Laboratory and Field Performance Tests

It has been seen that the new sets are superior in volume and in minimizing the disturbances of noise. The frequency measurements just discussed have indicated marked superiority also in the quality of reproduction.

One measure of the effect of this reduced distortion is by means of the articulation test. Such tests have shown that for a typical telephone system equipped with the new telephone sets, 95 per cent of the letter sounds spoken into the transmitter are correctly understood by the listener. With air transmission, 98 or 99 per cent of letter sounds are correctly understood. The difference is almost entirely due to the broader frequency band transmitted by the air path.

With the final designs of the new sets, tests have been made by the methods developed for determining "effective" transmission.[2] The results of these tests have shown that the new sets under the conditions of actual service provide a marked advance in transmission performance. The average total transmitting and receiving gain is about 15 db on the effective basis of transmission, as compared to sets of the sidetone type used with the older type of instruments.[13]

Conclusion

In general it appears that the notable transmission improvement which has been achieved in the design of the new telephone sets, in their freedom from distortion, higher effective volume, lower sidetone, and general convenience in use, makes possible a closer approach to the ease of a direct conversation than has hitherto been possible commercially.

Undoubtedly, further improvements in station transmission performance will, as in the past, be forthcoming with advances in the technique of design and manufacture, and in further knowledge of the requirements of the problem.

References

1. "Instruments for the New Telephone Sets," W. C. Jones, this issue of the *Bell System Technical Journal.*
2. "Rating the Transmission Performance of Telephone Circuits," W. H. Martin, *Bell System Technical Journal,* Vol. X, January 1931, pp. 116–131.
3. "Scientific Research Applied to the Telephone Transmitter and Receiver," E. H. Colpitts, *Bell System Technical Journal,* Vol. XVI, July 1937, pp. 251–274.
4. "Speech Power and Its Measurements," L. J. Sivian, *Bell System Technical Journal,* Vol. VIII, October 1929, pp. 646–661.
5. "Transmission Circuits for Telephonic Communication," K. S. Johnson, D. Van Nostrand Co., New York, 4th Printing, 1929, p. 105.
6. "Maximum Output Networks for Telephone Substation and Repeater Circuits," G. A. Campbell and R. M. Foster, *A. I. E. E. Transactions,* Vol. XXXIX, Pt. 1, pp. 231–280 (1920) (disc., pp. 280–290).
7. "An Explanation of the Common Battery Anti-sidetone Subscriber Set," C. O. Gibbon, *Bell System Technical Journal,* Vol. XVII, April 1938, pp. 245–257.
8. "A System of Effective Transmission Data for Rating Telephone Circuits," F. W. McKown and J. W. Emling, *Bell System Technical Journal,* Vol. XII, July 1933, pp. 331–346.
9. "Trend in Design of Telephone Transmitters and Receivers," W. H. Martin and W. F. Davidson, *Bell System Technical Journal,* Vol. IX, October 1930, pp. 622–627.
10. "Transmitted Frequency Range for Telephone Message Circuits," W. H. Martin, *Bell System Technical Journal,* Vol. IX, July 1930, pp. 483–486.
11. "Transmitted Frequency Range for Circuits in Broad Band Systems," H. A. Affel, *Bell System Technical Journal,* Vol. XVI, October 1937, pp. 487–492.
12. "A Voice and Ear for Telephone Measurements," A. H. Inglis, C. H. G. Gray and R. T. Jenkins, *Bell System Technical Journal,* Vol. XI, April 1932, pp. 293–317.
13. "Development of a Handset for Telephone Stations," W. C. Jones and A. H. Inglis, *Bell System Technical Journal,* Vol. XI, April 1932, pp. 245–263.

Spectrochemical Analysis * in Communication Research

By BEVERLY L. CLARKE and A. E. RUEHLE

The development of spectroscopy is traced through Newton, Fraunhofer and Kirchoff to Hartley, Pollock and Leonard, and De Gramont who, in the period 1880–1920, applied the spectroscope to chemical analysis. It is shown that modern quantitative spectrochemical analysis began with the comparison standard method of Meggers in 1922, and developments since that time are discussed. The organization and functioning of the spectrochemical unit of the analytical group at the Bell Telephone Laboratories is described, and a number of examples of applications to telephone problems are given.

The foundation of spectroscopy may be traced back to the year 1666 when Sir Isaac Newton discovered that sunlight is composed of the several colors, and that it could be separated into its components by refraction with a prism. Newton failed to note the discontinuous nature of the solar spectrum and it was reserved for Fraunhofer, more than a century later, to investigate the absorption lines and to point out their importance. Fraunhofer noticed that the *D* line absorption doublet in the solar spectrum was identical in position with the bright line doublet observed in a flame fed with sodium chloride. Finally in 1859 Kirchhoff formulated the modern concept of the composition of the sun based on the observed absorption lines in the solar spectrum.

As a result of Kirchhoff's theory much attention was at once turned to the examination and mapping of the emission spectra of terrestrial substances and testing for their presence in the solar atmosphere. Bunsen and Kirchhoff proved the presence of many terrestrial elements in the sun. Lockyer actually discovered one element, helium, in the sun almost thirty years before its discovery on the earth. Thus did qualitative spectral analysis enjoy a brilliant beginning.

In the years that followed Kirchhoff's work the spectroscope was properly credited with many triumphs in the field of physics, but applications to chemical analysis were extended very little beyond simple qualitative detection of the elements. In 1882, however, Hartley performed quantitative analyses by determining at what concentrations in solutions the various spectral lines of metals would disappear.

* Any process of chemical analysis by means of the emission spectrum will be called "spectrochemical analysis" in this paper. Absorption and fluorescence methods have also been used but these will not be considered here.

For example, in a 1 per cent solution of silver the spark spectrum contains forty-five lines, under the conditions of Hartley's experiments. In a 0.1 per cent solution of silver twenty-five lines remain; in a 0.01 per cent solution only nine lines remain; and at 0.001 per cent but one of these persists. Pollock and Leonard, following Hartley, also used solutions with an improved sparking cell and extended this method of quantitative analysis to most of the common metallic elements. Almost simultaneously De Gramont began his monumental work of determining the persistent and ultimate lines (*raies ultimes*) of all the known elements under the conditions of a condensed spark discharge.

In 1922 Meggers, Kiess and Stimson [1] published the first paper on what might be called the modern method of spectrochemical analysis. Their departure from previous methods was that they did not rely solely upon the presence or absence of certain lines as a measure of the amount of an element present in the sample. Taking cognizance of the fact that the concentration at which a line will just be visible depends upon a number of other factors besides the nature of the element, another means of standardization was adopted as follows: A graded series of standard samples was prepared with known amounts of impurity in the same base material as the alloy to be analyzed. This series of standards was sparked under the same conditions as the alloy in question and all spectra recorded on the same plate. This practically eliminated all variables of development and plate sensitivity and reduced the variables of excitation materially. The impurity content was determined by a visual comparison of the spectra of samples and standards.

The method of Meggers, Kiess and Stimson, known as the comparison standard method, is still used extensively. It is subject, however, to limitations as to reliability, sensitivity, and precision. It is not always possible to make up homogeneous alloys for standards, and since only a minute amount of sample is excited by the spark it may be difficult to excite a representative sample. This can be obviated by using solutions instead of alloys. Spark excitation is limited in its sensitivity and it is frequently necessary to analyze for smaller amounts than can be detected in the spark spectrum. The use of arc excitation usually overcomes this difficulty and furthermore the arc is somewhat more suitable for work with solutions, especially with small volumes. Visual comparison of sample with standard is limited in precision. In recent years, however, the use of the densitometer (*vide infra*) or of the logarithmic sector has increased the precision of comparison so that it is possible to make spectrochemical analyses with a precision of better

[1] Meggers, Kiess and Stimson, *U. S. Bureau of Standards Paper* No. 444 (1922).

than ± 5 per cent of the amount determined. In some cases a precision of ± 1–2 per cent has been claimed.

Having traced briefly the development of spectrochemical analysis, we shall now describe the functions of a modern spectrochemical laboratory and illustrate these functions by examples from the experience of the Bell Telephone Laboratories. At the outset it should be pointed out that this laboratory is confronted with an exceptionally wide variety of samples to be analyzed, due to the large number of materials used in the Bell System. Other spectrochemical laboratories, generally speaking, place a different emphasis on the various functions to be described. Furthermore, we work in cooperation with a resourceful and skillful analytical laboratory and a well organized microchemical laboratory,[2] both of which have developed methods for some materials covering the usual range of the spectrograph. As a result our functions have been limited in some directions but extended in others, so that again our experience may be somewhat different from that of other laboratories. On the other hand, the work of this laboratory illustrates the flexibility of the spectrochemical method by demonstrating that it can sometimes be even more useful in collaboration with other methods than in competition with them.

1. Qualitative Analysis for Metallic Impurities

This is probably the most common type of service performed. The spectrograph is admirably suited for such analyses since, although it utilizes only a small amount of sample, it can detect almost infinitesimal quantities of the metals and most of the metalloids. Furthermore, a complete qualitative analysis can be run in a few hours—if the spectrum is simple or the sample relatively pure one hour is usually sufficient. Thus in a relatively short time and using only a small portion of a completely unknown specimen a guide can be furnished for quantitative analysis by spectrochemical or other methods. As we shall see in the following section, it is possible also to obtain a fair idea of the quantities of each metal present from the same spectrogram.

Numerous examples could be cited but it will be sufficient to state that all "general unknowns" are first subjected to a spectrochemical qualitative examination before further analysis is attempted, and to give two typical cases. A silver contact was behaving abnormally in tarnish tests and was submitted for analysis. A qualitative test revealed that the silver contained thallium. Subsequent checks on the spools of silver wire from which contacts had been made showed

[2] Clarke and Hermance, "Microchemical and Special Methods of Analysis in Communication Research," *Bell Sys. Tech. Jour.*, *15*, 483 (1936).

that the wire on one spool also contained thallium and this wire was rejected for further use as contact material. Total time spent: three hours.

A lot of zinc straps for supporting aerial cables was rejected in the field because the straps were "brittle," i.e., when flexed they would crack rather than bend freely. Between areas which cracked there were areas which bent easily and showed no signs of breaking. Spectrochemical examination of the zinc at the point of fracture revealed the presence of mercury, while this element was absent in the flexible areas (Fig. 1). Time spent: two hours.

2. Quantitative Analysis by Estimation

The first question usually asked when a qualitative analysis shows the presence of an element is, "How much is present?" It is easy to to tell from a qualitative plate whether the element in question is present as a major component, a minor component, an impurity, or merely as a trace. This information is frequently the deciding factor as regards further analytical work. This is particularly important in diagnosing troubles in field complaints. Very often the spectrochemical evidence alone will settle the matter—if not it will almost invariably tell just what further analyses are necessary, thus promoting the greatest possible efficiency in reaching a final solution of the problem.

The method of making such a rough quantitative analysis is to rely upon the experience of the analyst in judging the percentage of the element necessary to give the observed density of the lines under the conditions used. While such a procedure is theoretically unsound, an experienced analyst can make a surprisingly accurate "guess" by this method as has been repeatedly demonstrated by subsequent quantitative analysis.

An excellent example of an application of this technique of estimation is the identification of contact materials. It is sometimes desirable to identify the material composing a contact without removing the apparatus from its mounting, without impairing the contact for further use, and sometimes without more than momentarily interrupting its operation. The method used is to rub the surface of the contact with a small piece of fine, especially pure abrasive paper and burn the paper in a graphite arc, photographing the resulting spectrum. From the lines appearing in this spectrum which do not appear in the spectrum of the abrasive paper alone the elements composing the contact can be identified and their proportions estimated (Fig. 2).

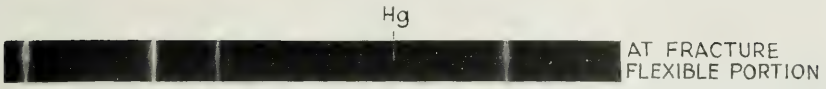


Fig. 1—Field complaint on failure of zinc straps for supporting aerial cables.

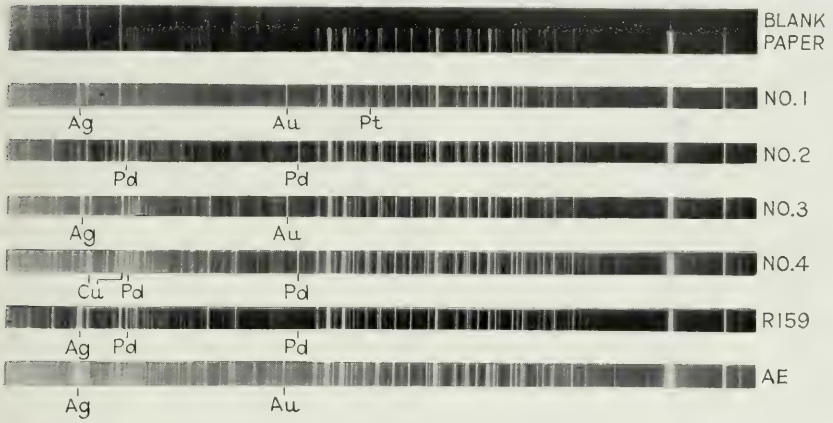


Fig. 2—Identification of contact alloys.

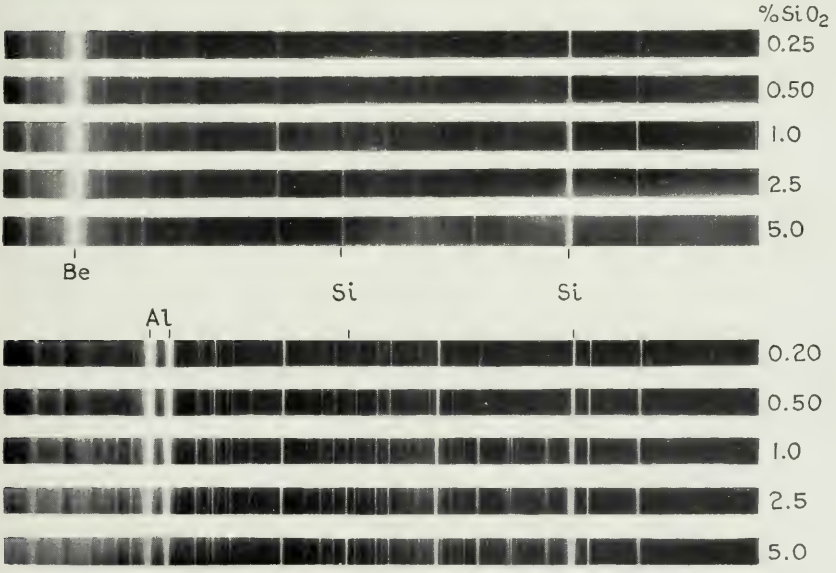


Fig. 3—Silica in beryllium oxide and aluminum oxide.

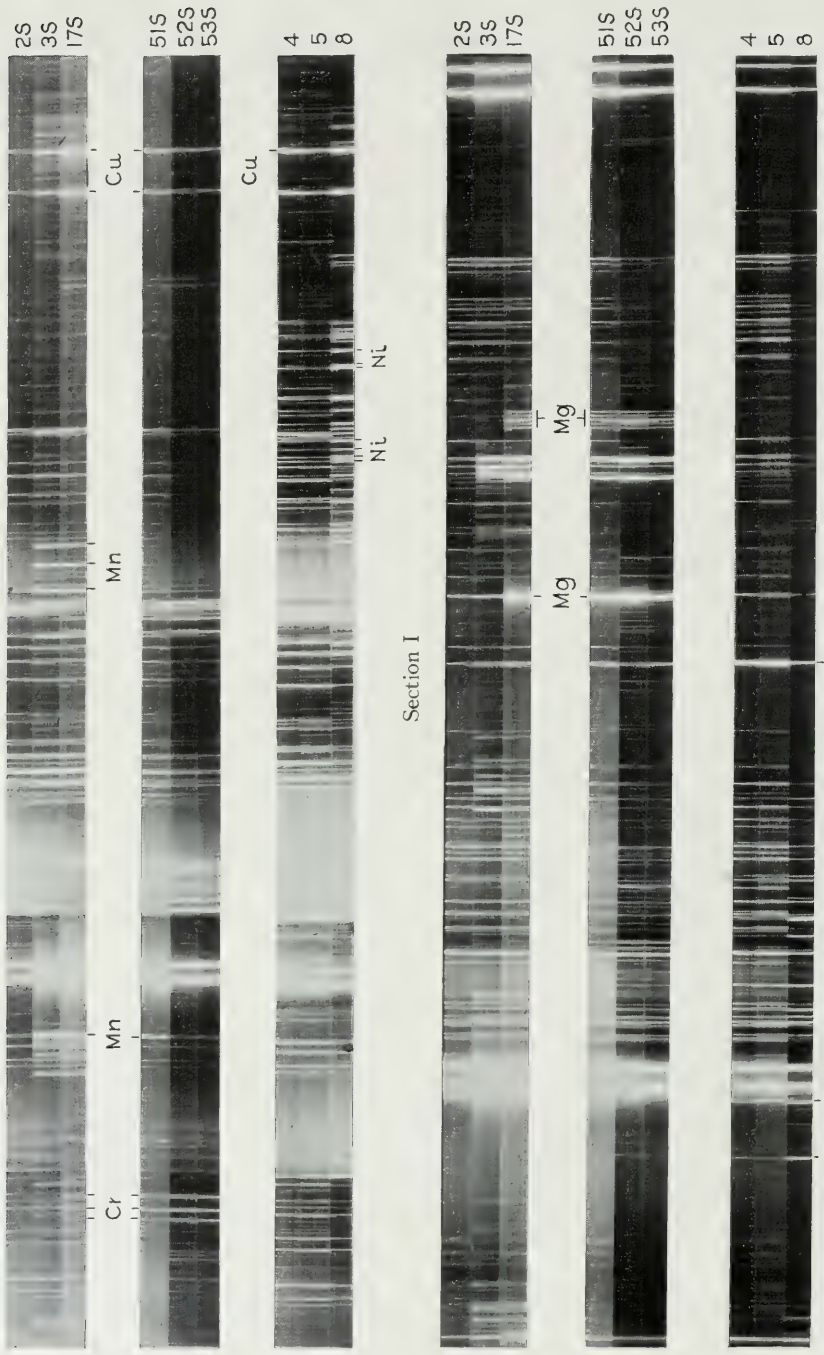


Fig. 4—Identification of aluminum alloys.

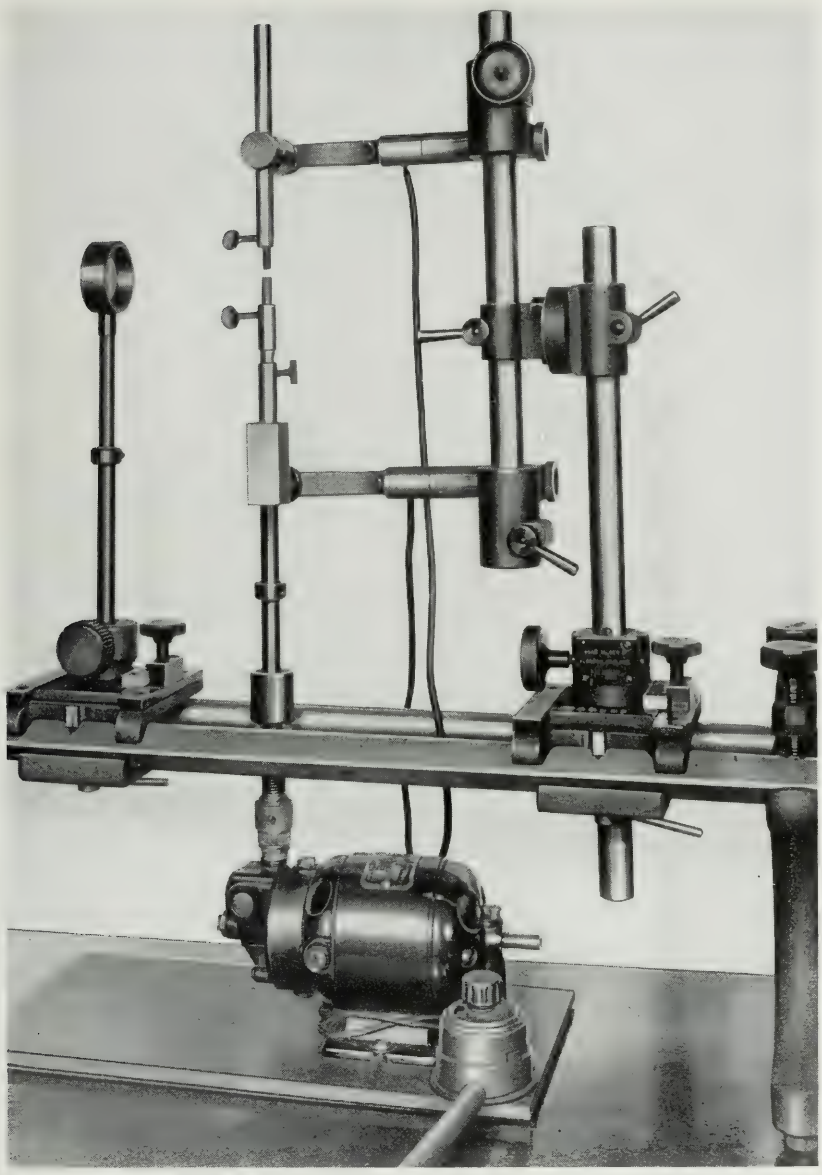


Fig. 5—Rotating electrode assembly adapted to De Gramont arc and spark stand.

Fig. 6—Step sector disk used for plate calibration.

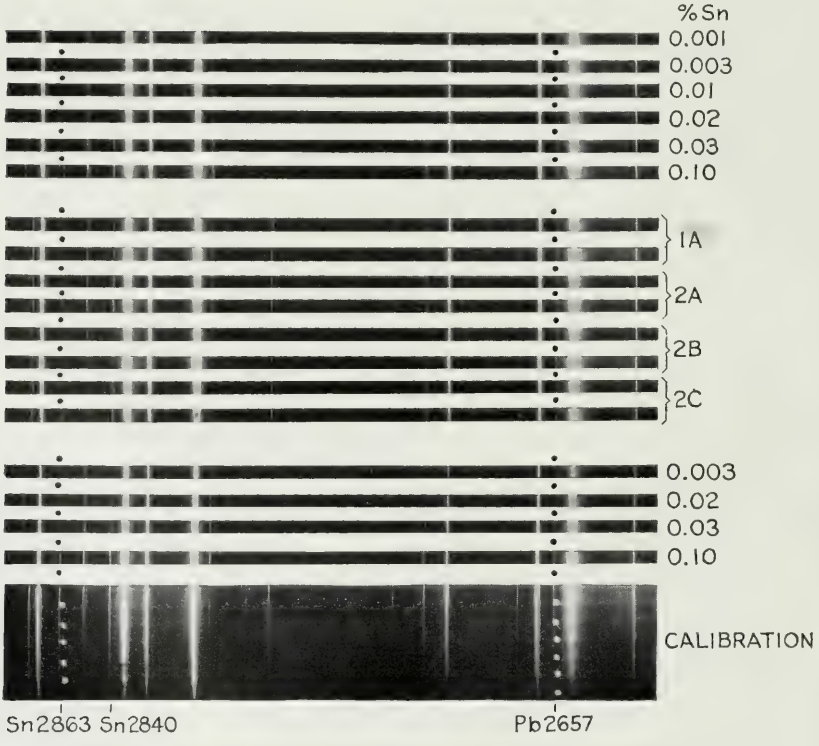


Fig. 7—Analysis of lead-antimony cable sheath for tin. An application of the internal standard method with plate calibration by the step sector.

Another example of this sort of analysis is the identification of aluminum alloys. It is a not uncommon occurrence for an alloy supposed to be of a certain composition to exhibit physical properties which would not be expected of such an alloy. Frequently it is important, therefore, to check the composition of the alloy and still preserve it for other tests. Now in the Bell System there are nine commonly used aluminum alloys, each of which contains different proportions of copper, manganese, magnesium, silicon, chromium and nickel. By examination of the spectrum obtained from a few grains of filings the analyst can easily tell by estimating the amounts of these minor components which of the nine alloys comprises the sample in question (Fig. 4).

3. Quantitative Analysis With High Speed and Moderate Precision

The greatest advantages of the spectrochemical method are speed, sensitivity, and flexibility of application. In many cases if a new material can be analyzed quickly only a moderate degree of precision is necessary. To fulfill such conditions an application of the method has been developed [3] which can be used for almost any material that can be dissolved in common solvents. The sample is first analyzed by method (1) or (2) above to find its approximate composition if that is not already known. Then a concentrated solution of the sample is prepared.

Another solution is prepared from pure specimens of the base materials present in the sample. To portions of this pure solution known amounts of the impurities present in the sample are added so that standard solutions are available for comparison with the sample. Spectra of the sample solution and the standard solutions are photographed on the same plate, using arc excitation on graphite electrodes, and the percentage of each impurity is estimated by comparison of line intensities. The method is capable of a precision of better than $\pm$ 10 per cent of the amount determined and can frequently be applied with little modification in cases where chemical methods would require considerable development work before they could be used. Furthermore the total sample needed for analysis is always very much less than is needed for a chemical analysis.

Since this is the quantitative method most widely applied in this laboratory numerous examples could be cited. To mention a few: aluminum, arsenic, tin and zinc in lead-base alloys, zinc in tin-base alloys, zinc in aluminum-base alloys, and cadmium, lead, tin, copper,

[3] Nitchie, C. C., *Indus. & Engg. Chem. (Anal. Ed.)*, *1*, 1 (1929).

iron, and magnesium in zinc-base alloys have been determined by this method.

4. Routine Analysis

In certain cases the spectrochemical method is superior to wet methods for routine quantitative analysis. In such cases dependence is placed on a set of solid standards whose composition is known, a set of standard solutions as in the foregoing section, or a standard working curve prepared by plotting the logarithm of the relative intensity of impurity lines as compared to lines of the base metal against the logarithm of the concentration of the impurity. In the latter case recourse is had to a photometric means of determining relative line intensity. The most common methods are the logarithmic sector method and the densitometer method (see next section). Both require considerable work to prepare the standard curve, but if many samples are to be analyzed periodically a saving in time results. The precision in such cases may be better than $\pm$ 5 per cent of the amount determined.

A few cases follow in which the spectrochemical method has proved superior for routine uses: (1) the analysis of zinc-base alloys, where it is more rapid; (2) the determination of zinc in tin-base alloys and in aluminum-base alloys, where no chemical or microchemical methods are available for the smaller amounts; (3) the determination of silicon in beryllium and aluminum oxides, where it is much more rapid (Fig. 3); (4) the determination of tin in lead-antimony cable sheath where it is more rapid; and (5) the determination of magnesium in nickel alloys where it is more economical of time and sample.

5. Quantitative Analysis with High Precision

Of recent years much of the published research on spectrochemical analysis has been concerned with attempts to improve the precision of the method. Applications of the densitometer and of the logarithmic sector have achieved something in this direction. Both of these devices have been used with some success in this laboratory. We have also developed a rotary electrode assembly (Fig. 5) which achieves a fairly steady mean position of the arc by causing it to vibrate rapidly about a point on the optical axis rather than to wander in a random manner over the electrode surface. The apparatus is applicable to the analysis of solutions in the graphite arc and to the analysis of alloys which can themselves be used as electrodes. Salts and other loose powders of course cannot be rotated at high speeds (600 R.P.M. is recommended).

The highest precision has been claimed by Duffendack and his

co-workers,[4] who use a "stepped diaphragm" to calibrate each plate with an intensity pattern in order to restrict their density measurements to the straight-line portion of the characteristic curve of the plate or to correct for deviations from the straight line portion. In this laboratory a step sector (Fig. 6) has recently been used to put the intensity pattern on each plate and the characteristic curve (Fig. 8) determined from this, using a photoelectric densitometer built in this laboratory. It is obvious from this curve that the measurement

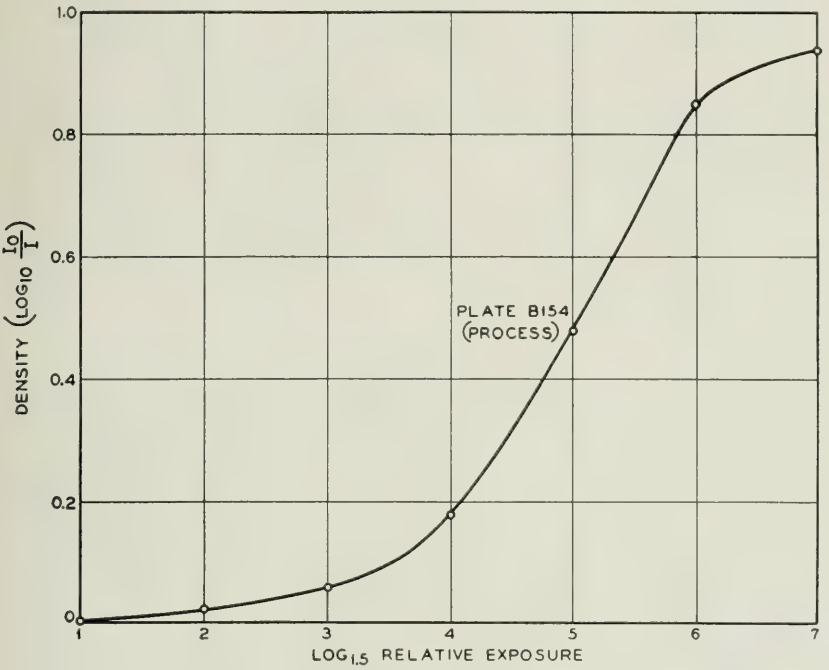


Fig. 8—A typical characteristic curve (exposure = intensity × time).

of the ratio of intensities of a given pair of lines varies with the densities of the lines as recorded on the plate. In this way errors of considerable magnitude may be introduced unless a calibration is made of the response of the plate to various levels of intensity of light of the wavelengths concerned. Results to date have shown an increased precision in all cases where plate calibration has been used, and a precision of better than ± 5 per cent of the amount determined is not uncommon. Examples from our experience include tin in lead-antimony cable sheath and other lead alloys (Fig. 7), magnesium in

[4] Duffendack, Wolfe, and Smith, *Indus. & Engg. Chem. (Anal. Ed.)*, *5*, 226 (1933) and others.

nickel filaments, and strontium-barium ratios in vacuum tube filament coatings.

6. Quantitative Analysis of Minute Traces

It has often been pointed out that spectrochemical methods can be used to assist classical and micromethods by establishing the purity of precipitates, checking the completeness of separation, demonstrating the presence or absence of interfering elements, etc. It has only infrequently been realized, however, that extremely minute concentrations of impurities can be determined by a combination of the two methods. In many cases a chemical separation can be made in which all of the impurities are removed from the bulk of the base material (it is not essential that *all* of the base material be separated as it is in chemical analyses) and can then be concentrated to fall within the spectrochemical range. In this way impurities can be determined by, say, method (3) above which could not be detected by direct excitation of the original sample. Scrupulous cleanliness and the carrying through of a reagents blank are, of course, essential when dealing with such small quantities.

7. As a Research Tool

Many times the spectrochemical method can be applied as a probe in obtaining important clues in a research project. This is particularly true when two presumably identical materials show unexpected differences in behavior. The research man may use nominally pure materials which are never really pure and thus introduce unsuspected impurities into his system which vary with different lots of materials. The spectrograph readily shows up differences in composition of the final product and thus often indicates the beneficial or detrimental effect of impurities.

In certain types of research on thin layers, the spectrochemical method is the only one of sufficient sensitivity to be used for quantitative analysis of the layers. Thus in thermionics it has been possible to measure the amounts of material on tungsten corresponding to changes in work function of the surface. Many other applications to thin films are possible and some are being worked on in this laboratory.

Thus it will be seen that chemical analysis by the emission spectrum is a powerful and versatile tool which enjoys and deserves an increasingly important role in communication research.

Acknowledgment

We are indebted to Mr. E. K. Jaycox for some of the method development as well as several of the spectrograms reproduced here.

High Speed Motion Picture Photography

By W. HERRIOTT

A motion picture camera used in taking 4000 pictures per second is described. Applications of high speed motion picture photography to a variety of problems associated with the design of telephone apparatus are given. The resulting pictures in "slow motion" permit convenient and accurate analysis of space-time relationships of mechanical parts in motion otherwise too rapid to be perceived because of their transient nature. This work has related to the development of relays, switches, clutches, ringers, dials, coin collector mechanisms, contact conditions, materials testing, etc. It has been applied to studies of noise reduction in mechanisms and to research problems associated with the production of speech by the vocal cords.

IN 1874 the French astronomer Jenssen pioneered in the use of motion picture photography as a visual aid to the study of a scientific problem. Considerable uncertainty then existed in the value of the earth's distance from the sun and Jenssen employed a camera capable of taking 48 pictures in 70 seconds during the transit of the planet Venus across the sun's disc. He hoped that errors of observation would be substantially less than those inherent in visual observations of this rare and important phenomenon but the results were disappointing due to certain photographic difficulties characteristic of the materials at his disposal. He recognized, however, the value of a series of rapidly taken photographs in making evident to the eye changes in the appearance and position of objects which would not otherwise be perceived because of their transient nature.

E. Muybridge was able to demonstrate in 1878 high speed motion pictures of animals in movement. Jenssen was content to examine his pictures singly under the microscope using the individual photographic images only for record purposes. Muybridge, however, made use of a simple viewing device in which the different pictures secured from a battery of cameras were viewed consecutively when mounted on the inner surface of a rotating cylinder provided with viewing slits. Again we have the application of photography to the discernment of transient movement not otherwise perceptible to the unaided eye. In the following year Muybridge was able to demonstrate the projection of motion pictures onto a viewing screen.

Many workers took up these pioneering experiments and improved devices were developed representing a continuous advance in the motion picture art until today we have a large industry applying the knowledge gathered by these workers to the educational and entertainment field. Paralleling commercial development of the motion picture there has been a continuous advance in its application to scientific and engineering problems, one phase of which relates to high speed motion picture photography.

Amateur and professional motion pictures are taken and projected at the rate of 16 or 24 frames or pictures per second. If pictures are taken at the rate of 48 per second and projected onto a viewing screen at the rate of 24 per second a magnification of the time axis by a factor of two occurs. The visual impression secured will be that of the same occurrence taking twice as long. If 480 pictures are taken per second and projected at the rate of 24 per second, the time magnification is 20. It is this magnification of the time axis which characterizes the picture as a "high speed motion picture." Cameras have been developed which will, under highly specialized conditions, extend the time axis by a factor of 2000 or 3000 times, although such phenomenal speeds of taking impose serious restrictions upon the nature of problems on which they may be used.

Motion pictures are usually made in a camera of the so-called "intermittent" type which refers to an intermittent motion given to the film by the film driving mechanism. An intermittent motion is employed in order that the film may be stationary during the brief exposure portion of the operating cycle after which it is rapidly accelerated and moved to an adjoining section for the next exposure. Mechanical difficulties limit the speed of operation of intermittent film moving mechanisms to a maximum taking rate of approximately 200 pictures per second. High speed motion pictures offering a magnification of the time axis by a factor of 10 can, therefore, be secured with specially constructed intermittent cameras. Higher speeds require the abandonment of the intermittent mechanism and the use of a non-intermittent or continuous film drive mechanism together with means for securing the sharp images required.

If film is moved continuously past an exposure aperture in which lies a stationary image of an object, obviously only streaks will result. Some means must be employed either to illuminate the object brightly for a sufficiently short length of time to avoid blurring of the photographic image or some device must be incorporated in the camera which will cause the image formed by the camera lens to be sharply focused on the film and to move in the direction of film travel with

the same velocity. Cameras operating on both principles have been developed and are used in scientific and engineering research. A large part of the work now being done by other workers in this field is with taking speeds extending from a few hundred pictures per second to 2,000 or 2,500 pictures per second.

Cameras of the first type, that is those in which the film is driven continuously and in which the object is brightly illuminated for a short length of time (usually of the order of 2 to 10 microseconds) are relatively simple in construction. Provision is made for the reel of unexposed film which may be 25, 50 or 100 feet in length and have either 16 millimeter or 35 millimeter width. The film may be guided past a fixed exposure aperture or around a rotating drum or toothed sprocket from which it passes to the take-up reel. Power may be supplied through an electric motor either to the take-up reel or to the toothed sprocket. If a stationary exposure aperture is used, two drive sprockets, placed one above and one below the gate, may be employed. Periodic flashing of the illuminant may be secured by the use of a commutator actuated by the camera mechanism.

The second type of camera above referred to employs an optical intermittent to produce the image movement required to avoid blurring when the film is continuously driven past the exposure aperture. A variety of optical intermittents has been developed employing either lenses, mirrors, prisms, or a plane parallel glass plate or block. Regardless of choice of means, the optical intermittent serves to produce a series of rapidly moving images which move with the film velocity. Of these available methods perhaps the simplest is the plane parallel glass plate or block which, by reason of its thickness, deviates or displaces an inclined ray in a manner nearly proportional to rotation about a chosen axis.

At Bell Telephone Laboratories a high speed camera has been developed which normally operates at a taking speed of 4,000 pictures per second. This camera employs optical compensation of the type in which a cube of glass is rotated at a high rate of speed (60,000 r.p.m. for 4,000 pictures per second) between the camera lens and the sprocket as shown in Fig. 1. The compensator cube has four polished faces, each parallel to its axis of rotation and parallel to the axis of the film sprocket. One picture is taken for each quarter revolution of the compensator. The index of refraction of the glass and the dimensions of the cube are chosen to cause correct movement of the image as the film is continuously advanced past the exposure area of the sprocket. The cube rotates in the direction of the arrow which is opposite to that of the sprocket. Downward movement of the image

formed by the camera lens results from change in refraction of light rays at opposite faces of the cube as the cube rotates. When the cube is in the position shown at "*A*" an upward displacement of the image to the point "*a*" results. Rotation of the cube in the clockwise direction diminishes the amount of displacement with the result that a downward image movement takes place which is synchronized with

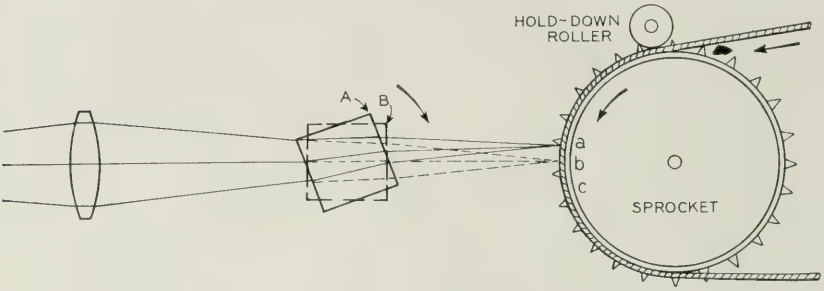


Fig. 1—Schematic arrangement of high speed camera.

the film movement. With the compensator cube in position as shown at "*B*," that is, its entrant and emergent faces perpendicular to the optical axis, no vertical deviation or displacement of the image results and the image falls at the point "*b*" on the film sprocket. Further rotation of the cube causes the image to move downward to the point "*c*" where the exposure is terminated and the next adjacent face of the cube comes into play. In this manner a succession of images is laid down frame by frame at a high rate of speed, each elemental area of the film having received exposure during a substantial part of the rotation cycle. The duration of each exposure is controlled by the film speed and by the angular height of a fixed aperture in front of each of the four faces of the cube. The film sprocket is directly driven from the motor shaft. Spur gears are employed to drive the optical compensator. A separate motor is employed to drive the take-up reel. 16 millimeter film in hundred foot lengths is used.

A finder is provided which permits viewing the image on the film as projected upon a ground glass screen mounted on the hinged door of the camera. Lenses of various focal lengths are interchangeable on the front of the camera. The camera is mounted upon a substantial tripod and is readily portable. Figure 2 shows the exterior of the Bell Telephone Laboratories high speed camera. Figure 3 shows the interior of this camera where the location of the film spools and main drive sprocket are shown. Figure 4 shows the camera with its two motors and portable lighting units of the type developed for

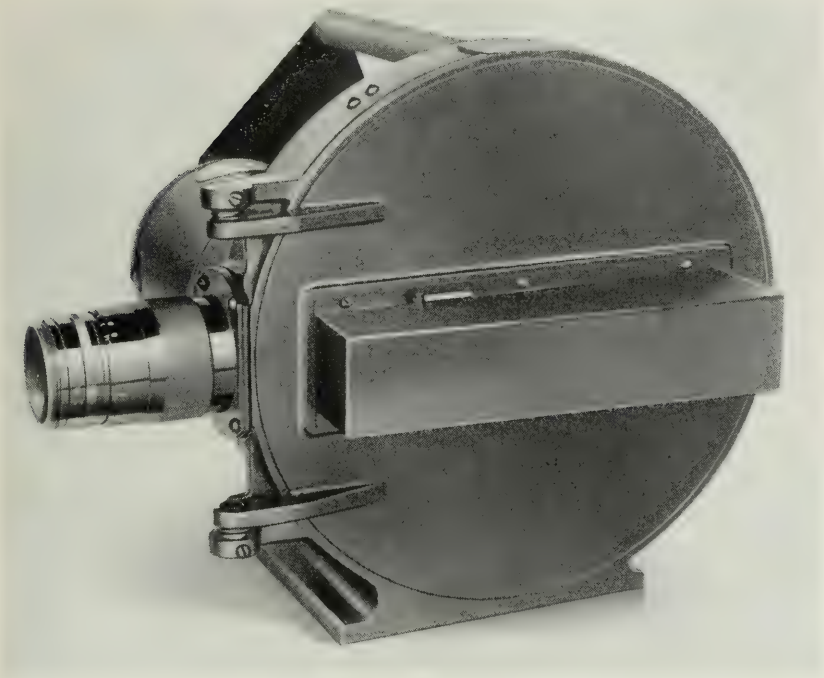


Fig. 2—High speed camera used in taking 4000 pictures per second.

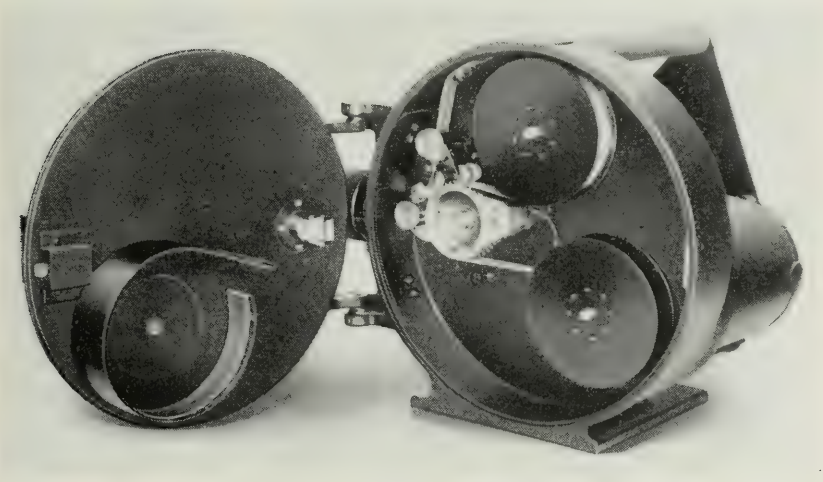


Fig. 3—Interior view of camera.

use in high speed photography. The camera has been adapted to use standard amateur 16 millimeter motion picture film and the commercial processing service of the film supplier is used. High-speed motion pictures in color are readily made, utilizing available films and processing.

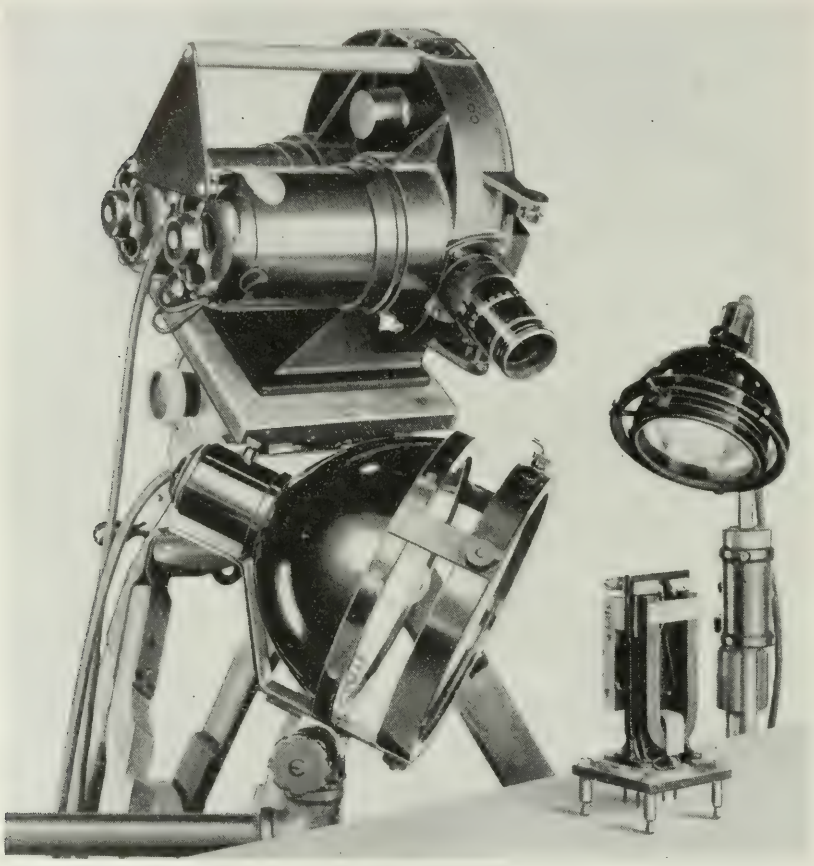


Fig. 4—Camera with portable lighting units used for high speed work.

The optical intermittent type of camera has certain advantages over other types. The unit is self-contained and is independent of lighting equipment. It can be made relatively light in weight and highly portable. Further, it can be applied to the study of self-luminous sources as required in the study of arc lamps, lamp burnouts, fuse burnouts and other problems where the phenomena under investigation are self-luminous.

High speed motion picture photography is finding extensive application in industry where it is applied to "motion analysis" of a variety of manufacturing operations and to problems associated with the design and performance of machinery. A particular value of this method of motion analysis lies in the convenience and accuracy with which space-time relationships of moving parts can be determined. Individual frames can be enlarged as photographic paper prints or projected onto a lined screen; in either case, the movement of parts can be measured with a high degree of accuracy. Extensive application is made in automotive and aeronautical engineering to the study of problems associated with fuel combustion, vibration in motors, airflow around structures and to propeller design and performance. It is coming into use in the fields of biology and medicine, especially in the study of muscular and nervous reactions. It has been applied to microphotography in the photographing of biological specimens at magnifications of 500 to 700 times at a taking rate of 1000 pictures per second.

In high speed photography one of the principal problems has been the securing of adequate illumination of the subject. Exposure times are of the order of one twenty-thousandth second or less for taking speeds of 4000 frames per second. Artificial illumination must be used and arc lamps and tungsten incandescent lamps are employed. Special projection bulbs are secured which are greatly over-voltaged during the few moments required in the taking of a picture. The lamps are designed for short life and to operate as close to the melting point of tungsten as is practicable. Two lamps are operated in series during setting up and adjustment of the apparatus and full voltage applied to both lamps at the moment of taking by the use of series-parallel switching arrangements. The lamps are housed in specially designed lighting units employing high aperture lenses or mirrors which serve to image the source directly on the object at a desired magnification. Provision is made for the reduction of heat by the use of water cells of suitable thickness. Excessive heat in the image will frequently cause distortion in delicate apparatus which must be avoided. One or more lighting units may be employed depending upon the size of the object being photographed, the taking speed, the lens aperture and the film speed. Brightnesses of the order of 10,000 to several hundred thousand foot candles are frequently employed.

At Bell Telephone Laboratories, high speed photography has been applied to a wide variety of problems associated with design and performance of telephone apparatus. Pictures have been taken of standard equipment and of experimental equipment in course of

development. Much valuable data not otherwise obtainable have been secured regarding functioning of telephone equipment under use conditions. This work has related to relays, switches, clutches, ringers, dials, coin collector mechanisms, contact conditions, etc. It has been applied to the testing of materials in connection with impact testing, stress analysis and bending moment. It has also been applied to reduction of noise in apparatus, particularly in machines of the type widely employed for accounting purposes, coin counting, typewriter mechanisms, and other high speed operating mechanisms where it was desired to analyze and remedy certain noise conditions. It has found particular use in fundamental research associated with the production of speech by the vocal organs. High speed motion pictures have been made of the vocal cords in action employing taking speeds of 4000 frames per second. These pictures are particularly valuable to teachers of speech and music and to the medical profession. Such pictures have, of course, great value to engineers working on problems associated with the transmission of speech over telephone circuits.

A service in high speed photography is available to engineers of the Laboratories as a visual aid to their study of problems associated with the design, manufacture and performance of telephone apparatus. The high degree of portability which has been achieved in both the camera and lighting equipment lends itself well to extensive application of this service to engineers. Figures A to L show series of individual frames illustrating a variety of problems to which this service has been applied.

High speed motion pictures of a variety of subjects have been taken in color. Color pictures have been made of vocal cords and of photo-elastic effects revealed in transparent materials under polarized light. Such pictures are valuable in studying stress and impact conditions as affecting design of equipment. Stereoscopic high speed motion pictures have been made both in color and in black and white.

High speed motion picture photography is finding increased application to a variety of problems associated with the work of the Laboratories. It is believed that more extended applications of its use will follow.

TYPICAL HIGH SPEED MOTION PICTURES

A—High speed photographs of an experimental model of the step-by-step switch showing the wiper rising to the cut-in position and overthrowing on the first contact of its associated bank. Motion pictures of this type disclosed in great detail movements of the wiper involved in the operation of this switch. These movements are so rapid that but little is learned from a visual examination of the switch while in operation. 4000 pictures per second.

B—These pictures show normal action of the vertical pawl in the step-by-step switch. Again, high speed motion pictures show much that cannot be gained

from visual examination. In this case, the pictures revealed whether or not the action of the pawl and ratchet was satisfactory under extreme operating conditions. 2000 pictures per second.

C—To study the conditions under which handset breakage might occur, apparatus was developed in which breakage takes place under controlled conditions. High speed photographs revealed distortions in the handset which under certain test conditions resulted in breakage at the moment of impact. Information relating to the structural strength of the handset was secured from measurements made on many individual pictures of this series. The pictures shown illustrate breakage of an experimental 3-piece E type handset after falling from a height of 5 feet and striking a rigid steel bar shown at the bottom of each picture. 4000 pictures per second.

D—Information gained from a study of high speed pictures of the type represented by C resulted in strengthening of the reenforcement of the handle at certain points which reduced the possibility of breakage under normal use conditions. 4000 pictures per second.

E—High speed photography is applied extensively to the study of explosion of gases in motors, to ballistic problems associated with the explosion of gun powder and to other rapid phenomena of a self luminous type. This series of pictures shows the melting and burning of fuse wire under heavy current conditions. 20-ampere fuse wire is shown during burn-out on direct short. The violence and extent of the action are well shown in these pictures. 4000 pictures per second.

F—Certain normally isotropic transparent materials become birefringent when examined in a stressed condition under polarized light. Extended use of this effect is made in the study of stress distributions in engineering structures and in models of mechanical parts. High speed photography is now applied to these photoelastic effects exhibited in a glyptol sample under impact stress condition. This series of pictures shows impact testing of an unnotched glyptol specimen in plane polarized light. 300 pictures per second.

G—Poor contact conditions in relays may give rise to improper circuit operation. High speed motion pictures have been useful in the study of contact chatter in relays and other similar devices. This series of pictures shows normal operation of contacts. 2000 pictures per second.

H—This series of pictures exhibits contact chatter. In the first picture of this series the movable contact spring is shown contacting the left fixed contacts. In the second and third pictures, the movable contact spring has been drawn against the right hand fixed contacts. At this point the current is cut off and the movable contact springs return to normal as shown in the sixth picture. Chatter occurs at this point with the movable spring returning to make contact with the stationary contacts shown at the right. Two cycles of chatter condition are shown. 2000 pictures per second.

I—This series of pictures shows the No. 14 teletypewriter locking arm lever and cam during overthrow which gives rise to noisy operation. They illustrate a typical source of objectionable noise in apparatus of this type. Excessive clearance between the cam and the locking arm lever is shown which results in impact noise on the return of the locking arm lever. 1800 pictures per second.

J—This series of pictures shows a modified No. 14 teletypewriter locking arm lever and cam in which the overthrow has been eliminated with subsequent reduction in noise. It can be seen that the lever arm now closely follows the contour of the cam. 1800 pictures per second.

K—A knowledge of the fundamentals of speech and hearing is important to designers of telephone apparatus. High speed motion picture photography has been applied to problems associated with the production of speech by the vocal mechanism. The pictures show vocal cords vibrating in production of speech sound at a frequency of 120 cycles per second. Pictures of this type offer a unique and practical means of securing much useful information relating to the production of speech. 4000 pictures per second.

L—At L is shown the action of the clapper striking one gong of an experimental 20-cycle ringer. This picture revealed more strokes of the clapper per second of operation than was desired. This condition resulted in a peculiar acoustic effect, readily explained from this series of pictures. 2000 pictures per second.

A
B
C
D

E F G H

I
J
K
L

Bibliography

1. "Motion Picture Camera Taking 3200 Pictures per Second," C. F. Jenkins, *Trans. Soc. Motion Picture Engineers*, No. 17, Oct. 1923, p. 77.
2. "The Chronoteine Camera," C. F. Jenkins, *Jour. S. A. E.*, No. 22, Feb. 1928, p. 200–2; *Trans. Soc. Motion Picture Engineers*, No. 25, Sept. 1926, p. 25.
3. "New High Speed Kinematographic Camera," T. Suhara, *Proc. Imp. Acad.*, Tokio, No. 5, Oct. 1929, p. 334–7.
4. "New Ultra-Speed Kinematographic Camera," T. Suhara, *Report No. 60, Tokio Univ. Aeronautical Res. Inst.*, May 1930, p. 187–94.
5. "A Non-intermittent High-Speed 16 M.M. Camera," F. E. Tuttle, *Jour. Soc. Motion Picture Engineers*, No. 21, Dec. 1933, p. 474.
6. "The N. A. C. A. Photographic Apparatus for Studying Fuel Spray From Oil Engine Injection Valves and Test Results From Several Researches," E. G. Beardsley, *N. A. C. A. Tech. Report No. 247*, 1927.
7. "The N. A. C. A. Apparatus For Studying The Formation and Combustion of Fuel Sprays and the Results From Preliminary Tests," A. M. Rothrock, *N. A. C. A. Tech. Report No. 429*, 1932.
8. "Stroboscopic and Slow Motion Pictures by Means of Intermittent Light," H. E. Edgerton, *Jour. Soc. Motion Picture Engineers*, No. 3, 1932, p. 356.
9. "The Mercury Arc as an Actinic Stroboscopic Light Source," H. E. Edgerton and K. J. Germeshausen, *Rev. of Scientific Instruments*, Oct. 1932, p. 535.
10. "High Speed Motion Pictures," H. E. Edgerton, *Electrical Engineering*, Feb. 1935, Vol. 54, p. 149–153.
11. "Precision Timing of Athletic and other Sporting Events," C. H. Fetter and H. M. Stoller, *Electrical Engineering*, June 1933, Vol. 52. p. 386–391.
12. "Time Microscope for High Speed Machines," J. P. Maxfield, *Modern Packaging*, July 1936, Vol. 9, No. 11.
13. "High Speed Motion Pictures of the Vocal Cords," W. Herriott and D. W. Farnsworth, *Radiography and Clinical Photography*, June 1938, Vol 14, No. 3, pp. 26–27.

An Optical Harmonic Analyzer *

By H. C. MONTGOMERY

An instrument which makes a Fourier Series Analysis of a function by optical means has recently been completed. The function to be analyzed is supplied in the form of a variation in the density or in the width of the transparent portion of a photographic film. The analysis is performed by a direct evaluation of the integrals which form the coefficients in a Fourier Series, and the results are theoretically exact in the sense that the measurement of each harmonic is independent of the other harmonics which may be present in the function. The operation of the instrument is largely automatic, and is rapid enough so that 30 harmonics can be measured in about a minute and a half.[1]

A PERIODIC function can be represented for all values of the variable by a Fourier Series. A function which is not periodic can be so represented between any finite limits, although the series may be entirely unlike the function beyond these limits. If a function is approximately periodic, the Fourier Series representing adjacent portions of it will generally be approximately alike.

Although in general an infinite number of terms is required to represent a function exactly, it is common experience that a great many functions of practical interest can be closely approximated by a series of from ten to thirty terms.[2]

Principle of Operation

The principle of this analyzer was suggested by E. C. Wente of these Laboratories.[3] It may be outlined as follows.

The Fourier Series expansion of a function is given by either of the following equivalent expressions.[4]

* Presented at Meeting of Acoustical Society of America, Washington, D. C., May 3, 1938.

[1] For comparison, analysis to 30 harmonics on the Henrici type instrument requires five or six hours. A resonance analyzer, such as the vibrating reed type, can complete an analysis in a few seconds, but the phases will not be given, and if the function is provided in graphical form it must be converted into an electrical or acoustic wave form repeated enough times for the resonant elements to reach a steady state response.

[2] A description of a number of the more important methods of harmonic analysis, together with a bibliography, is contained in "Sound Analysis," H. H. Hall, *Jour. Acous. Soc. Amer.*, vol. 8, pp. 257–262, April 1937.

[3] U. S. Patent No. 2,098,326.

[4] The expressions in this form apply when the fundamental period is 2π. There is no loss of generality, as the scale of abscissa can always be so chosen as to conform

$$f(x) = a_0 + \sum_1^\infty a_n \cos nx + \sum_1^\infty b_n \sin nx \tag{1}$$

$$= c_0 + \sum_1^\infty c_n \cos (nx - \phi_n). \tag{2}$$

Comparison of (1) and (2) gives the following relations between the coefficients in the two forms of the expression:

$$a_n = c_n \cos \phi_n, \qquad b_n = c_n \sin \phi_n. \tag{3}$$

$$c_n^2 = a_n^2 + b_n^2, \qquad \phi_n = \tan^{-1} \frac{b_n}{a_n}. \tag{4}$$

The form (2) giving amplitude and phase angle of the harmonics is generally more useful, but most methods of analysis give the coefficients in form (1) necessitating the computation of the amplitude and phase angle from the relations (4). One of the advantages of the optical analyzer is that it will give either set of coefficients directly.

The coefficients in the Fourier Series can be determined from the following expressions: [5]

$$a_n = \frac{1}{\pi}\int_0^{2\pi} f(x) \cos nx\, dx, \qquad b_n = \frac{1}{\pi}\int_0^{2\pi} f(x) \sin nx\, dx. \tag{5}$$

$$c_n = \frac{1}{\pi}\int_0^{2\pi} f(x) \cos (nx - \phi_n) dx, \quad \int_0^{2\pi} f(x) \sin (nx - \phi_n) dx = 0. \tag{6}$$

$$a_0 = c_0 = \frac{1}{2\pi}\int_0^{2\pi} f(x) dx. \tag{7}$$

We will now describe two methods by which a function can be represented on a photographic film. In a *variable area record* the film, which is elsewhere opaque, contains a transparent portion whose width at any point is proportional to the function. Such a record is shown in the upper part of Fig. 1. In a *variable density record* the

to this requirement. In fact this selection of a proper scale corresponds directly to a necessary adjustment of the analyzer.

[5] The expressions given in (6) can be derived from the more familiar expressions (5) as follows. From (3)

$$a_n \cos \phi_n + b_n \sin \phi_n = c_n(\cos^2 \phi_n + \sin^2 \phi_n) = c_n,$$
$$b_n \cos \phi_n - a_n \sin \phi_n = 0.$$

Substituting values of a_n and b_n given by (5) in these expressions leads at once to the expressions (6).

function is represented by gradations in the density of the film such that the light transmission at any point is proportional to the function, the density being uniform in a direction perpendicular to the axis. Such a record is shown schematically in the lower part of Fig. 1. With either type of representation of a function $g(x)$, it will be seen that the amount of light transmitted through a narrow vertical strip of width dx is proportional to $g(x)dx$. If two or more such records are superimposed, the light transmitted through all of them will be proportional to the product of the recorded functions, provided not more than one of the records is of the variable area type.

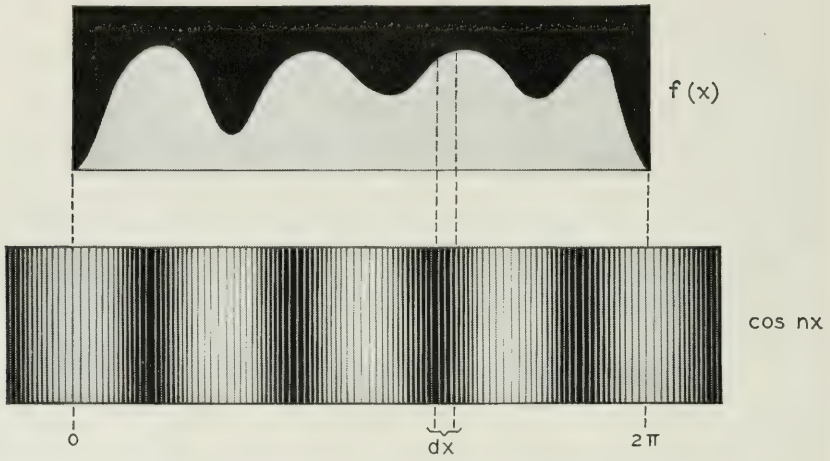


Fig. 1—Representation of $f(x)$ and cos nx on film.

The determination of a_n and b_n is now very straight-forward. Suppose we have $f(x)$ recorded on one film and cos nx on another. For illustration we will assume that $f(x)$ is recorded in variable area and cos nx in variable density, as shown in Fig. 1, although the only necessary requirement is that they shall not both be variable area. If the two films are superimposed, the amount of light transmitted through both of them between the limits zero and 2π is just the first integral in (5) and hence proportional to a_n. Similarly, if the cosine screen is moved a quarter wave-length along the axis it becomes sin nx and we have at once b_n. If the cosine screen is moved a whole wave-length along the axis, the transmitted light will go through a maximum. It will be shown below that this maximum value is proportional to c_n, and the position of the cosine screen at which it occurs is ϕ_n.

One more matter needs to be considered before we write down the expressions which describe the operation of the analyzer. Since $f(x)$

and cos nx will in general have both negative and positive values they cannot be directly represented by the transmission of light, which is essentially positive. However, the addition of a constant to each function will eliminate this difficulty, and merely results in a constant in the measured amplitude, as shown below.

The optical transmission of the film on which $f(x)$ is recorded may be written

$$A + f(x),$$

where A is a constant large enough to make the expression positive for all values of x. Similarly, the transmission of the cosine screens may be written

$$B_n[1 + M_n \cos (nx - \theta)],$$

where θ is a parameter denoting the position of the cosine screen along the x-axis, M_n is a constant somewhat less than unity, known as the modulation of the record, and B_n is a constant which is seen to be the average optical transmission of the screen.

If one or both of these records is of the variable density type, the total transmission when they are superimposed will be

$$\begin{aligned} T &= \int_0^{2\pi} B_n[A + f(x)][1 + M_n \cos (nx - \theta)]dx \\ &= \int_0^{2\pi} AB_n dx + \int_0^{2\pi} B_n f(x)dx + \int_0^{2\pi} AB_n M_n \cos (nx - \theta)dx \\ &\qquad + \int_0^{2\pi} B_n M_n f(x) \cos [(nx - \phi_n) - (\theta - \phi_n)]dx, \\ &= 2\pi B_n(A + c_0) + \pi B_n M_n c_n \cos (\theta - \phi_n). \end{aligned} \tag{8}$$

To obtain a_n, we take the difference in T for $\theta = 0$ and $\theta = \pi$, which is seen to be

$$2\pi B_n M_n c_n \cos \phi_n = 2\pi B_n M_n a_n. \tag{9a}$$

Similarly, to obtain b_n, we make $\theta = \frac{\pi}{2}$ and $\theta = \frac{3\pi}{2}$, giving for the difference in T

$$2\pi B_n M_n c_n \sin \phi_n = 2\pi B_n M_n b_n. \tag{9b}$$

To obtain c_n and ϕ_n note that the maximum value of T occurs at $\theta = \phi_n$ and the minimum at $\theta = \phi_n + \pi$, which serves to determine ϕ_n. The difference between the maximum and minimum values of T is

$$2\pi B_n M_n c_n. \tag{10}$$

If the factors B_n and M_n can be made approximately constant for all the screens, the coefficients for either form of the Fourier Series are directly proportional to the change in the amount of transmitted light for specified pairs of positions of the cosine screens.

Description of the Instrument

The process which the analyzer is required to carry out consists of superimposing the function to be analyzed on a cosine screen and measuring the variation in the transmitted light when the cosine screen is moved along the x-axis. This is repeated with a different cosine screen for each harmonic which it is desired to measure.

A schematic diagram of the instrument is shown in Fig. 2. The film containing $f(x)$ is placed in a holder at A and strongly illuminated by

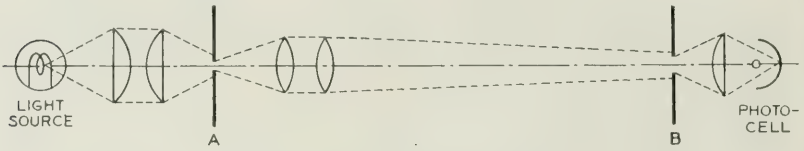


Fig. 2—Diagram of optical system.

an incandescent lamp and condensing lens. An enlarged image of $f(x)$ is formed at B on a window bounded by two knife edges .750 inch apart. Functions of different length are accommodated by adjusting the optical enlargement so that the image of the portion of $f(x)$ to be analyzed will just fill the window. The cosine screens slide in a track directly behind the window, and receive the image of $f(x)$. The transmitted light is collected by another lens and brought to a photocell.

A series of cams and levers is arranged to bring the cosine screens out of a drum shaped magazine in which they are stored into the optical path, give them the small motion required for analysis, and return them to the magazine, which is then rotated to bring the next screen into position. These operations are all automatic, and the attention of the operator is required only for the adjustment of the enlargement and focus and resetting of the cams at the beginning of each analysis. A photograph of the instrument is shown in Fig. 3.

The variations in the photocell output take place at the rate of about two cycles per second. These are recorded on a moving chart by an instrument similar to a high speed level recorder,[6] differing from it chiefly in having a linear instead of a logarithmic scale.

[6] "A High Speed Level Recorder," Wente, Bedell and Swartzel, *Jour. Acous. Soc. Amer.*, vol. 6, p. 121, January 1935.

The present instrument has been designed to take records of $f(x)$ which are from one-sixteenth to five-sixteenths of an inch long and no higher than their length. The focal length of the enlarging lens is 1.5 inches. The collecting lens is placed quite close to the cosine screens, and forms an image of the enlarging lens on the plate of the photocell. With this arrangement the patterns of both $f(x)$ and the cosine screens are well diffused on the photocell plate, so that surface variations in sensitivity of the plate are unimportant. The illumination is uniform across the field to ± 2 per cent.

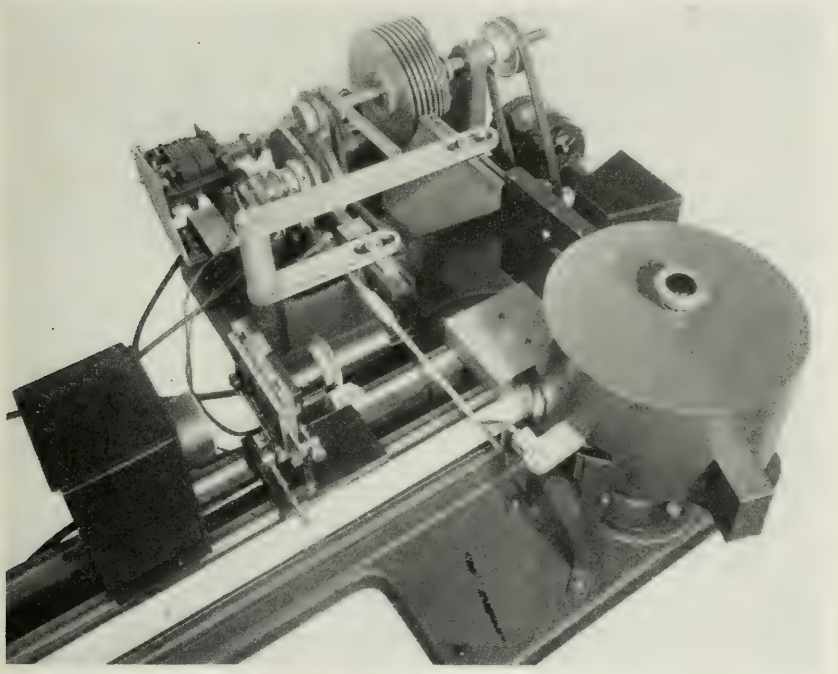


Fig. 3—The optical harmonic analyzer.

The cosine screens were made on photographic plates, by printing from variable density motion picture sound track negatives containing records of pure single frequencies. The pattern thus produced is about 1×2 inches. The increase in width of the track from about one-tenth of an inch in the negative to one inch on the plate was secured by making a "contact" print with negative and plate slightly separated, and moving the plate sideways under the negative while printing.

The important requirements for the screens are good wave form,

uniformity in modulation and average transmission, and accuracy in wave-length. In the present instrument it was found possible to keep the harmonic content of the screens down to 5 per cent. The modulation varied from 79 per cent to 94 per cent in different screens, and the average transmission from 20 per cent to 24 per cent. Variations in the wave-length of the screens amounted to about 1 per cent.[7]

It is convenient though not necessary to have good wave form in the screens. When readings are made in pairs as described above, the effects of even order harmonics in the cosine screens cancel out. Moreover, G. R. Stibitz has shown [8] that the cosine screens may have practically any wave form (not even necessarily periodic), and correction factors can be derived for them. The process of correction is rather cumbersome, however, since the correction for each harmonic is not a constant, but depends on other harmonics present in the function.

Uses of the Analyzer

This instrument was designed particularly to accommodate sound records on film as used in commercial motion picture work. However, functions from any other source can be analyzed equally well if they are reproduced with the proper dimensions on film. Provision is made for measurement of the first 30 harmonics. As stated above, the function must be between 1/16 and 5/16 inch in length and no higher than its length. At the speeds customarily used for recording sound on film this corresponds to a fundamental frequency of from 65 to 310 cycles per second, or 1950 to 9300 cycles per second for the 30th harmonic.

The smallest harmonic which the instrument will indicate is about 2 per cent of the peak which it can accommodate. In connection with this statement, it should be remembered, however, that for many functions the largest harmonic in the analysis is considerably less than the peak amplitude of the function, which reduces the effective amplitude range.

An interesting check on the operation of the analyzer can be obtained by making an analysis of a simple geometric wave form. For example, a single cycle of a saw-toothed wave can easily be formed by placing a straight edge obliquely across the sound track slot of the analyzer. Such a wave form is shown at the top of Fig. 4. It is known that this wave form can be resolved into a series of har-

[7] Since a 1 per cent error in wave-length amounts to an error of about one-third of a wave in the total length of a screen of the 30th order, errors of this magnitude are quite objectionable in the higher order screens, although unimportant in the low orders.

[8] Unpublished work.

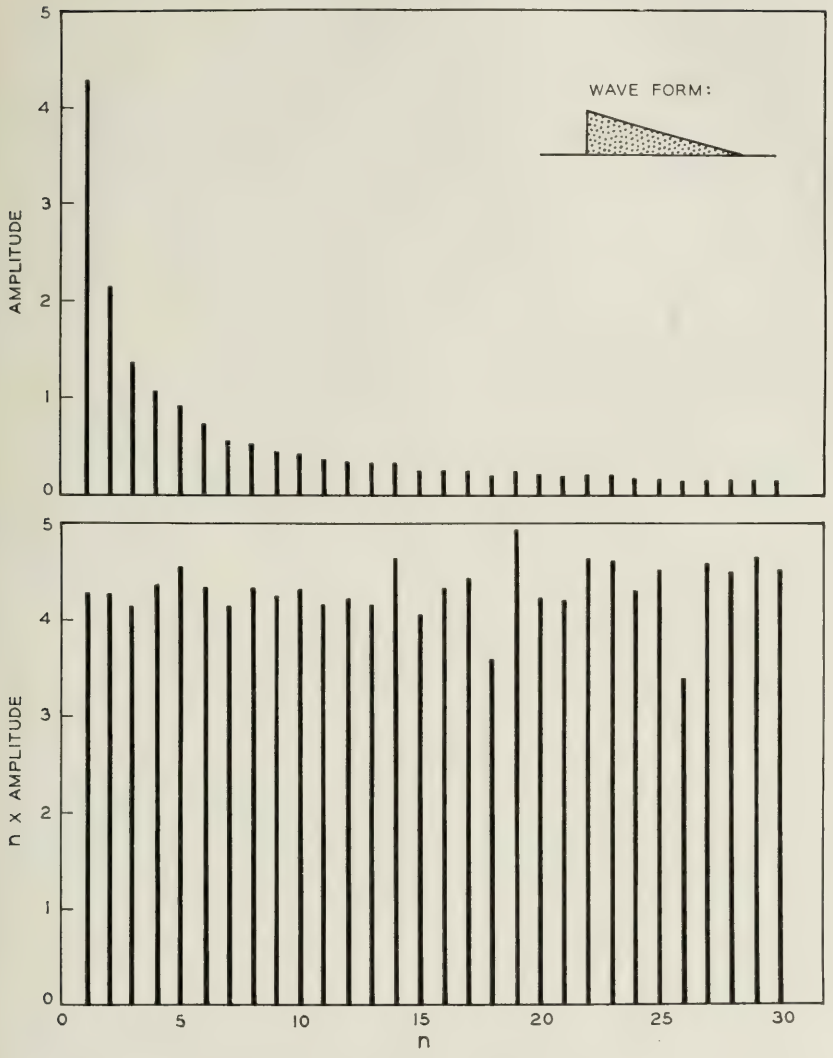


Fig. 4—Analysis of saw-toothed wave.

monics whose amplitude falls off as $1/n$ where n is the order of the harmonic. The values obtained with this analyzer are shown in the upper graph in the figure. In the lower graph each harmonic has been multiplied by n, which should make all the ordinates equal if the analysis were exactly correct.

The use of the analyzer for the sounds of speech is illustrated in Fig. 5, which shows the analyses of portions of two vowel sounds made

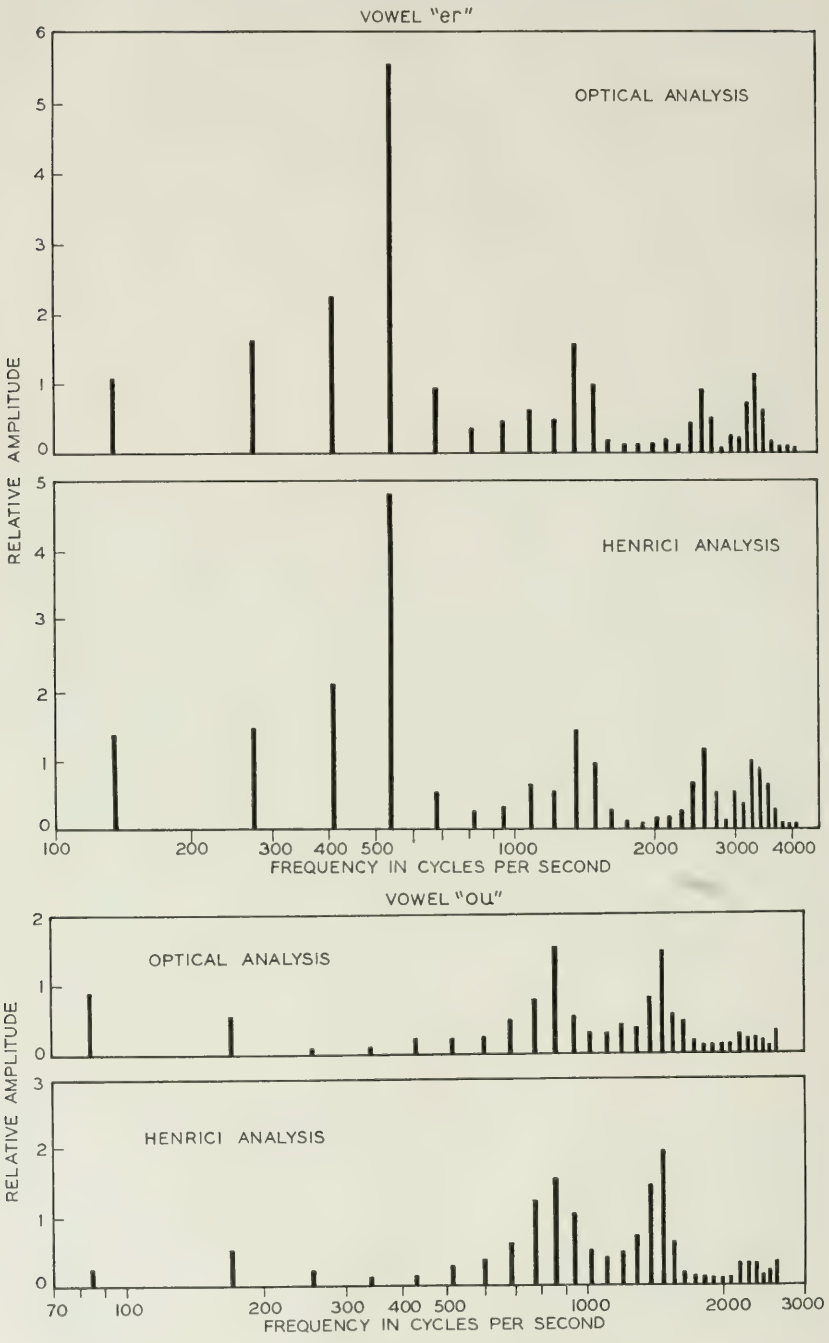


Fig. 5—Analyses of spoken vowel sounds.

with the optical analyzer. Each is compared with an analysis of the identical wave form made on a Henrici type analyzer at the State University of Iowa.[9] The first sound is a portion of the *er* in *father*. There is a very prominent fourth harmonic, indicating a strong resonance in the voice at 530 cycles. Other smaller peaks occur at 1400, 2650 and 3500 cycles. The second sound is a portion of the diphthong *ou* in *out*. It shows two peaks of about equal magnitude, with a suggestion of a third smaller one. The general features of the analyses by the two methods are seen to be in good agreement. A series of such analyses throughout the course of a spoken sound furnishes a fairly complete description of the changes in resonance, amplitude, and fundamental frequency which are taking place. Because of its high speed of operation and convenient application to records of speech on film, the present form of the optical analyzer is especially adapted to such a study of the characteristics of connected speech.

[9] "The Henrici Harmonic Analyzer," D. C. Miller, *Jour. Franklin Inst.*, vol. 185, pp. 285–322 (1916).

Magnetic Shielding of Transformers at Audio Frequencies

By W. G. GUSTAFSON

The first part of this article is a descriptive discussion of magnetic shielding in general. Formulae are then given for the calculation of shielding efficiency of cylindrical shells for steady and alternating magnetic fields. By means of these formulae the shielding efficiency for various types of cylindrical shields has been calculated for a steady magnetic field.

The second part of the article contains experimental information on various types of transformer shields. This information supplements the theory in connection with factors which would be very laborious to treat theoretically.

The theory and the experimental data are coordinated in such a manner that the shielding efficiency of a particular shield can be calculated with an accuracy which is sufficient for practical purposes.

IN connection with the development of repeaters for long distance telephone lines it is found that noise is introduced into the telephone lines due to magnetic pick-up by transformers and coils in the repeaters. This applies also to sound pictures equipment, public and private address systems, etc., where high gain amplifiers are used. The stray magnetic field causing this pick-up may be produced by neighboring generators, transformers, rectifiers and other power equipment. It may also be produced by other amplifier transformers and coils or by relays located in the vicinity of the disturbed coil. The intensity of the disturbing field may frequently be of the order of 0.1 oersted at the point of pick-up. However, a field intensity of the order of 0.02 oersted often causes objectionable noise and under extreme conditions values as low as 0.001 oersted may be undesirable. As the gain of the amplifier increases and the demand for good quality becomes greater, it becomes increasingly important to control magnetic pick-up. The limiting of this pick-up is in fact today one of the important problems to be considered in the design of high-gain amplifiers.

One method by which the magnetic pick-up can be decreased is by arranging the core structure and winding distribution of the transformer in such a way that the voltages induced by an external field are at least partially neutralized. In many cases, however, this

impairs other important characteristics of the transformer and is therefore undesirable.

Another method is by shielding the transformer from the disturbing magnetic field. It is the object of this paper to consider such shielding and to present some data in this connection that may be of general interest.

Theory

When a transformer is placed in an a.c. magnetic field, there will, in general, be a voltage induced in the windings. This voltage is proportional to the intensity of the magnetic field. Therefore, if the intensity of the magnetic field in the space occupied by the transformer is reduced, the induced voltage will be correspondingly reduced. This can be accomplished by enclosing the transformer in a case made of material which shields against magnetic flux. Let H_i be the intensity of the field inside the case and H_e the intensity of the field when the case is removed. The ratio H_e/H_i will then indicate the shielding efficiency of the case. Expressed in decibels:

$$\text{Shielding efficiency} = 20 \log_{10} H_e/H_i. \tag{1}$$

The shielding efficiency of the case depends primarily upon the permeability and conductivity of the material, and the mechanical construction of the case.

A high permeability material provides a magnetic path in the walls of the case of much less reluctance than the air space inside the case. The greater part of the flux will, therefore, follow the low reluctance path, and only a small part will enter the space inside the case. The higher the permeability is, the less the flux that will enter the space inside the case.* With a steady magnetic field all the shielding is due to this cause.

An alternating magnetic flux induces eddy currents in the material of the case as shown in Fig. 1. These eddy currents are a function of the conductivity and permeability of the material. They may increase or decrease the shielding efficiency of the case. That is, the eddy currents i_{e1} (Fig. 1), which are due to the component of the magnetic field perpendicular to a wall of the case, will set up a counter mmf. which will oppose flux entering the case. In a copper case, the shielding is primarily due to such eddy currents. On the other hand, the eddy currents i_{e2} (Fig. 1), which are due to the component of the field parallel to a wall of the case will set up a counter mmf.

* It is assumed here, of course, that the source of the magnetic flux is at some distance so that the amount of flux leaving the source is not appreciably affected by the case.

which will oppose the flux following the low reluctance path in the walls of the case, or what is the same thing, decrease the effective permeability of this path and will, in that way, decrease the shielding efficiency. In a case made of high permeability material the latter eddy currents, i_{e2}, obviously should be reduced as much as possible.

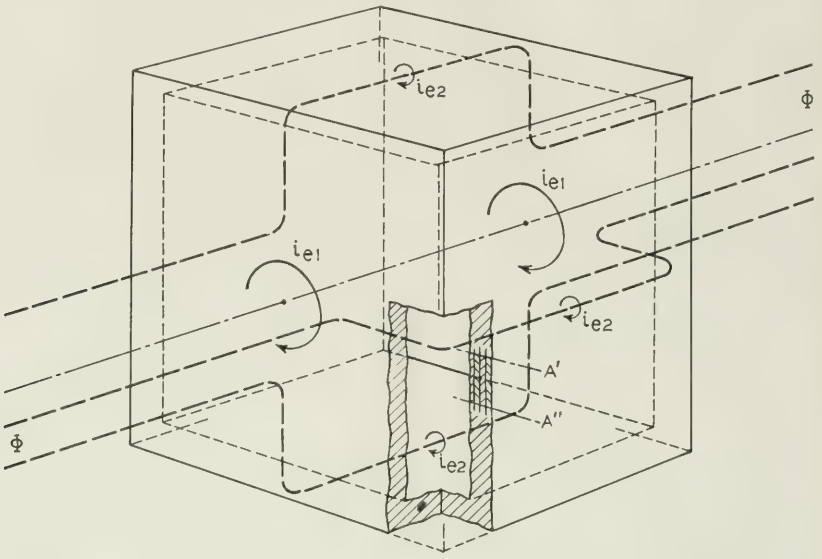


Fig. 1—Eddy currents produced in a shield by an alternating magnetic field.

If the resistivity of the material is increased, both sets of eddy currents will be decreased. If, however, the material is laminated with the sheets parallel to the wall of the case as indicated by section A'–A'' in Fig. 1, the undesired eddy currents will be reduced without affecting those which are beneficial.

The relative effectiveness of the low reluctance path and eddy currents in securing good shielding efficiency against magnetic fields depends mainly upon the frequency of the magnetic field. As a general rule, we can say that at low frequencies, the effect of the low reluctance path predominates, while the shielding effect of the eddy currents, i_{e1}, increases as the frequency increases.

This way of looking at the effect of permeability and conductivity in a magnetic shield is intended to be purely descriptive and probably would not be practicable for a mathematical treatment. It is, however, very suggestive to the design engineer.

As an illustration of the way in which the mechanical construction of the case may affect the shielding efficiency it is at once clear that

the ratio between the reluctance of the magnetic path in the walls around the case and the path through the interior of the case depends upon the size of the case. Also, any openings in the case will obviously affect its shielding efficiency.

When a magnetic transformer core is placed inside the case, the reluctance through the interior of the case will decrease and the shielding efficiency of the case will also decrease. It is therefore evident that with one type of magnetic core a case might have a different shielding efficiency than with another.

To obtain general mathematical relations between the shielding efficiency and the various factors mentioned above is very difficult. By making some simplifying assumptions, however, relations can be obtained that will be useful from a practical standpoint, although they will necessarily be somewhat limited in application.

A great deal of work on the shielding efficiency of shields constructed of different magnetic materials has been done by various investigators.* Except in a few of the more recent papers,† consideration has been restricted to a steady, uniform, magnetic field where no eddy currents are produced and where, therefore, the shielding is due entirely to the magnetic properties of the material. They have also usually limited themselves to spherical and cylindrical shields. The cylinders have been considered of infinite length with the direction of the disturbing magnetic field perpendicular to the axis of the cylinder. However, with cylinders of finite length they have found that for moderate shielding efficiencies, at points inside a cylinder at a distance from the end equal to its diameter the shielding efficiency is approximately that of an infinite cylinder. The ratio between the shielding efficiency of a cylinder and that of a sphere, the radii of the two, the permeability, the thickness and construction of the walls being the same, varies from approximately 4 : 3 in favor of the sphere for very thin shells to 9 : 8 in favor of the cylinder for very thick shells. This gives some idea of how the shape of the shield affects the shielding efficiency.

Investigations by the various investigators referred to above show that the shielding efficiency of two or more concentric cylinders or spheres may be vastly greater than that of one cylinder or sphere, the amount of magnetic material being the same. They have given mathematical relations between the shielding efficiency, the permeability and the mechanical dimensions of both spheres and cylinders. Although these relations have been derived for a steady magnetic field, they may also be applied with certain limitations to an alternating

* See Bibliography.

† The most important exceptions are articles No. 18, 19, 21, and 28 in the Bibliography.

magnetic field. The permeability used in this case will, of course, be the effective permeability for the particular conditions and frequency under consideration. These relations refer only to that portion of the shielding effect which is independent of the eddy currents i_{e1} (Fig. 1), and the total shielding effect will, in general, be somewhat greater.

In an article in the *Physical Review* of October, 1899, A. P. Wills considers the cases of three concentric cylinders and spheres. Due to the fact that spherical shields are less suited for our purpose I will give the equations for cylindrical shields only. Wills' formula for three cylinders for large values of permeability is given by the following equation:

$$\begin{aligned} g = 1/4\,\mu\,\{&(1 - q_1q_2q_3) + 1/16\,\mu^2 n_1 n_{12} n_2 n_{23} n_3 \\ &+ 1/4\,\mu[(n_1n_3 + n_1n_2 - n_1n_2n_3)n_{12} \\ &+ (n_1n_3 + n_2n_3 - n_1n_2n_3)n_{23} - n_1n_3n_{12}n_{23}]\} + 1. \end{aligned} \tag{2}$$

In this equation

$$g = \frac{H_e}{H_i}, \tag{3}$$

where H_e is the density of the magnetic field at a point P with the shield removed and H_i is the density of the magnetic field at that point when enclosed by the shield. μ is the permeability of the material at the frequency in question. We have

$$\begin{aligned} q_1 &= r_1^2/R_1^2, & n_1 &= 1 - q_1, \\ q_2 &= r_2^2/R_2^2, & n_2 &= 1 - q_2, \\ q_3 &= r_3^2/R_3^2, & n_3 &= 1 - q_3, \\ q_{12} &= R_1^2/r_2^2, & n_{12} &= 1 - q_{12}, \\ q_{23} &= R_2^2/r_3^2, & n_{23} &= 1 - q_{23}, \end{aligned} \tag{4}$$

where r_1, R_1, r_2, etc., are the various radii of the cylinders as shown by Fig. 2.

By making $q_3 = 1$ (or $n_3 = 0$), in (2), we get the relation for two concentric cylinders.

$$g = 1/4\,\mu(1 - q_1q_2 + 1/4\mu n_1n_2n_{12}) + 1. \tag{5}$$

By making $q_2 = 1$ (or $n_2 = 0$), (5) changes into an equation for one shell only

$$g = 1/4\,\mu(1 - q) + 1. \tag{6}$$

It has been shown (A. P. Wills, *Phys. Rev.*, vol. 24, page 243, February 1907) that for a shield of predetermined size, that is when the smallest and the largest radii (r_1 and R_3 in the case of three cylinders, Fig. 2) are specified, the radii of the surfaces of the successive

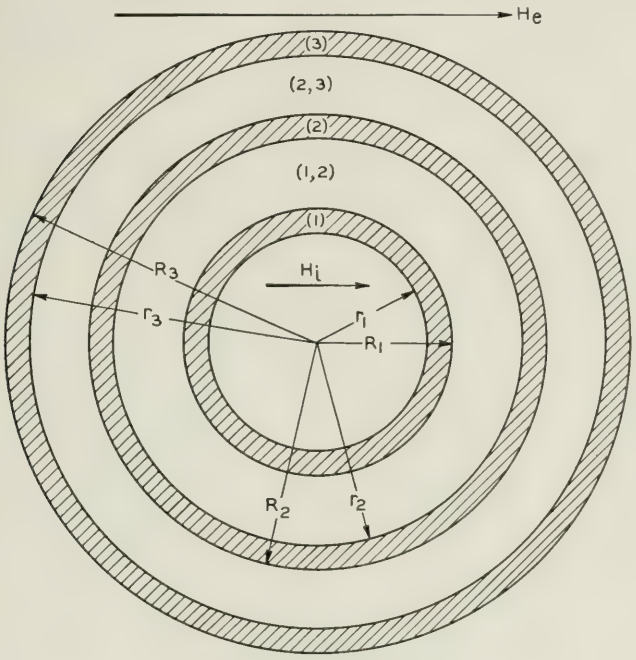


Fig. 2.

cylinders should be in a geometric progression to give the most efficient shield. That is, we should have $q_1 = q_2 = q_3 = q_{12} = q_{23} = q$. Equation (2) then becomes

$$g = 1/4\ \mu(1 - q^3 + 1/16\ \mu^2 n^5 + \mu n^3 + 3/4\ \mu n^4) + 1. \qquad (7)$$

For two cylinders we get

$$g = 1/4\ \mu(1 - q^2 + 1/4\ \mu n^3) + 1. \qquad (8)$$

In these equations, the following relations hold between r_1 and R_3 for (7) and r_1 and R_2 for (8)

$$R_3 = r_1/\sqrt{q^5}, \qquad R_2 = r_1/\sqrt{q^3}. \qquad (9)$$

The effect upon the shielding efficiency of varying s (Fig. 3) from

zero to P keeping $r_1/R_1 = r_2/R_2$ can be obtained by means of the following equation

$$g = 1/4\,\mu\left[1 - \frac{k^2}{q_{12}} + 1/4\,\mu\left(1 - \frac{k}{\sqrt{q_{12}}}\right)^2 n_{12}\right] + 1, \tag{10}$$

where

$$k = \frac{r_1}{R_2}.$$

If R_1/r_2 (that is $\sqrt{q_{12}}$) is varied from 1 to R_2/r_1 the desired result is obtained.

Assume in Fig. 3 the thickness of the two cylinders to be the same, that is, $R_1 - r_1 = R_2 - r_2$. The variation in shielding efficiency vs. s

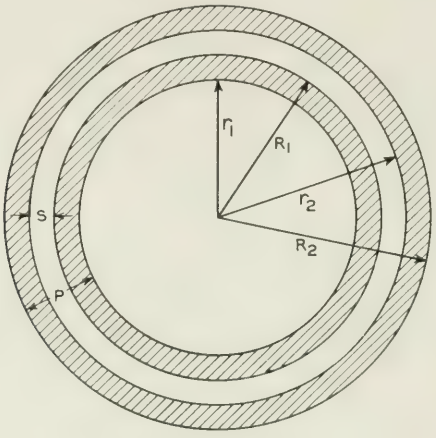


Fig. 3.

or $\sqrt{q_{12}}$, is then given by equation (5), with q_2 expressed in terms of q_{12} and q_1 as follows:

$$\sqrt{q_2} = \frac{1}{1 + \sqrt{q_{12}}\left[1 - \sqrt{q_1}\right]}. \tag{11}$$

In an article in the *Philosophical Magazine* of February 1933 L. V. King has developed relations for the shielding efficiency of spherical and cylindrical shells taking into account the effect of induced currents. The following equations for an infinitely long metallic cylinder have been picked from his paper. For a non-magnetic shell, the thickness of which is small compared to its radius, the shielding ratio, g, is given by

$$g = |\cosh(ka) + 1/2\,ka \sinh(kd)|, \tag{12}$$

where a is the radius and d the thickness in cm. k is given by

$$k = 2\pi\sqrt{\frac{f}{\rho \cdot 10^9}}(1 + i), \tag{13}$$

where f is frequency in cycles and ρ is resistivity in ohms per centimeter cube. At low frequencies (12) reduces to

$$g = \left| 1 + i\frac{4\pi^2 fad}{\rho} 10^{-9} \right|, \tag{14}$$

which is good up to about 10^5 cycles. The direction of the disturbing magnetic field in (12) and (14) has been assumed perpendicular to the axis of the cylinder.

Other formulæ which take into account both conductivity and permeability are also given in King's article. They are, however, rather complicated and require elaborate calculations.

Mr. S. A. Schelkunoff in an article in the October 1934 issue of the *Bell System Technical Journal* has derived formulæ which are comparatively simple although they take into account both conductivity and permeability. His treatment is quite different from that presented above and his results are expressed in terms of radial impedances. For an infinitely long cylindrical shield the diameter of which is large compared to the radial thickness of the shield the shielding efficiency, S, is given by

$$S = R + A. \tag{15}$$

In this formula R is the sum of the reflection losses at the surfaces of the shield. We have

$$R = \sum_{n=1}^{n} 20 \log_{10} \frac{|k_n + 1|^2}{4|k_n|} \text{ db}, \tag{16}$$

where k_n is the ratio of the radial impedance in the first medium to that in the second. That is,

$$k_n = \frac{Z_1}{Z_2}. \tag{17}$$

The radial impedance for a good dielectric is given by

$$Z = 2\pi f\mu\rho i \text{ ohms}. \tag{18}$$

For a metal

$$Z = \sqrt{\frac{2\pi f\mu i}{g}}. \tag{19}$$

In (18) and (19) f is the frequency in cycles, ρ is the radius in cms., g is the intrinsic conductance in mhos/cm., and μ is the intrinsic inductance in henries/cm.*

A in (16) is the sum of the attenuation losses in the successive shells. For any one shell

$$A = 8.686\alpha t \text{ db}, \tag{20}$$

where $\alpha = \sqrt{\pi g \mu f}$ and t is the radial thickness in cms.

Calculated Curves

By means of the equations (2) to (11) the shielding efficiency of various types of cylindrical shields has been calculated. The permea-

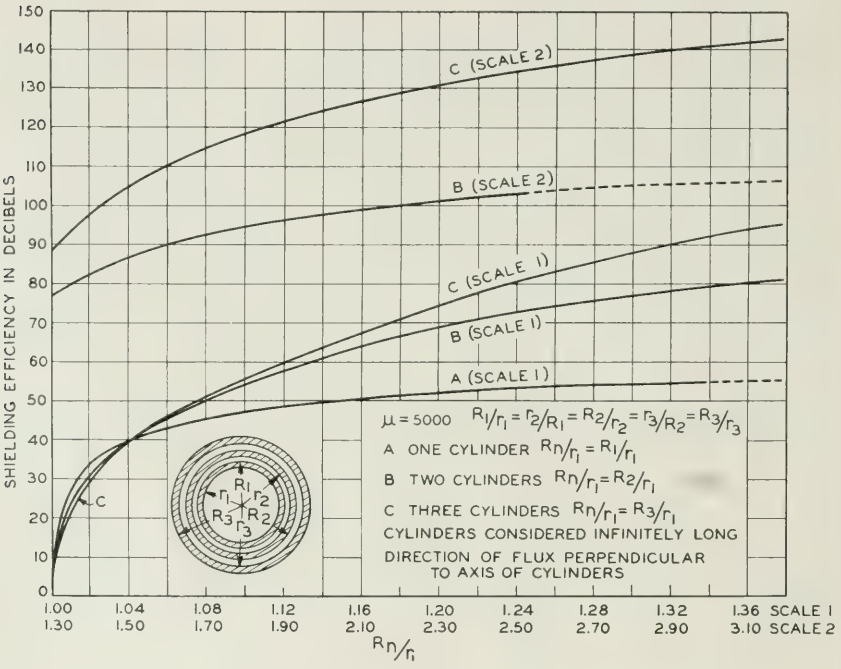

SHIELDING EFFICIENCY IN DECIBELS
C (SCALE 2)
B (SCALE 2)
C (SCALE 1)
B (SCALE 1)
A (SCALE 1)
μ = 5000 R1/r1 = r2/R1 = R2/r2 = r3/R2 = R3/r3
A ONE CYLINDER Rn/r1 = R1/r1
B TWO CYLINDERS Rn/r1 = R2/r1
C THREE CYLINDERS Rn/r1 = R3/r1
CYLINDERS CONSIDERED INFINITELY LONG
DIRECTION OF FLUX PERPENDICULAR TO AXIS OF CYLINDERS
Rn/r1
SCALE 1
SCALE 2

Fig. 4—Shielding efficiency of one, two, and three concentric cylinders (for zero frequency).

bility considered is 5000 which is readily obtainable at low frequencies and field strengths by means of permalloy.† The calculations are given in the form of curves in Figs. 4, 5 and 6.

* Thus in empty space (or dielectrics, approximately) $\mu = 4\pi 10^{-9}$ henries/cm. In general $\mu = 4\pi\mu_0 10^{-9}$ where μ_0 is the intrinsic permeability referred to empty space as unity.

† Arnold and Elmen, "Permalloy," *Journal Franklin Institute*, May, 1923, pp. 621–632.

The curves of Fig. 4 show the shielding efficiency of one cylinder and of two and three concentric cylinders with air space between. The shielding efficiency is given as a function of the ratio between the outside and inside radii of the shield, that is, R_n/r_1, where R_n is the outside radius of the outside cylinder and r_1 is the inside radius of the inside cylinder. These curves show the relative shielding efficiencies of 1, 2 and 3 cylinders and give numerical values for μ equal to 5000. The relative dimensions of the cylinders are such that the ratios between the inside and the outside radii of the cylinders and of the air spaces between the cylinders are in geometric progression. These curves show that when a very high shielding efficiency is desired it is not only advantageous but necessary to use two or more cylinders. Thus with one cylinder the maximum shielding efficiency that can be considered practical when $\mu = 5000$, is approximately 50 db. The maximum theoretical limit for one, two and three cylinders is 62 db, 124 db and 186 db respectively when $\mu = 5000$.

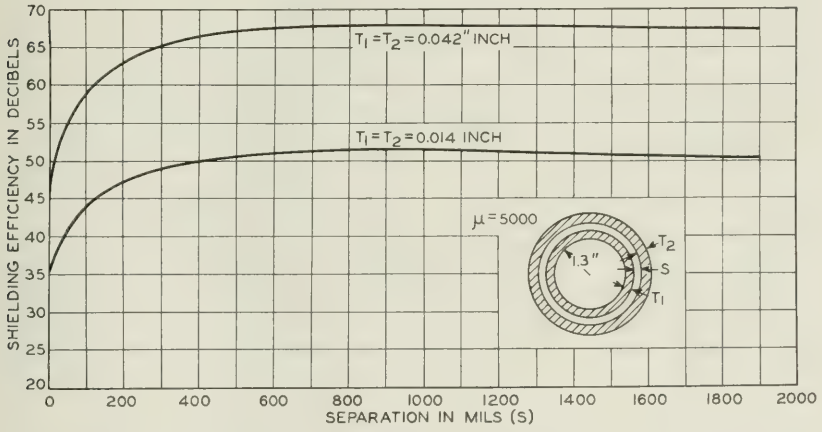


Fig. 5.—Shielding efficiency of two concentric cylinders versus the air-gap between them. Thickness of the wall of each cylinder kept constant. Zero frequency.

In Fig. 5 is shown the shielding efficiency of two concentric cylinders vs. the thickness of the air space between the cylinders. The thicknesses of the walls of the two cylinders are equal. Two thicknesses of the walls of the cylinders have been considered, namely, .014″ and .042″. An interesting fact is brought out by comparing the curves of Fig. 5 with curve "*B*" of Fig. 4. For example, when the air space is .042″, the upper curve of Fig. 5 gives a shielding efficiency of two cylinders with an air space between them, the thickness of which is the same as that of the cylinders. The ratio between the outside

radius and the inside radius is 1.097. The shielding efficiency of this combination is approximately the same as that of two cylinders having the same ratio R_n/r_1 as given in Fig. 4, curve "*B*." This shows that the condition that the radii of the cylinders should be in geometric progression is not very critical, at least, not for the value of R_n/r_1 under consideration.

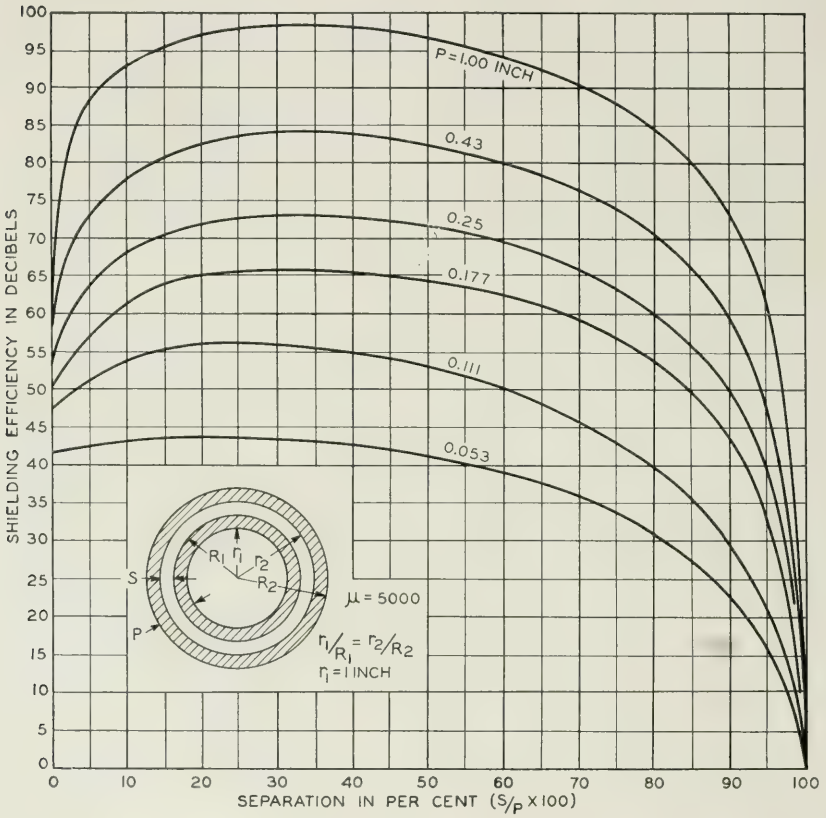


Fig. 6—Shielding efficiency of two concentric cylinders versus the air-gap between them. Overall thickness of the wall of the shield (P) kept constant. Zero frequency.

The curves of Fig. 6 also show the shielding efficiency of two concentric cylinders vs. the air-gap between them. In this case, however, the total thickness P of the wall of the double cylindrical shield is kept constant and the air-gap is increased from zero to P at the expense of the cylinders. When the air-gap is zero, the shielding efficiency is therefore that of a solid cylinder, the thickness of the

wall of which is P. When the air-gap is equal to P, the thickness of the two cylinders is zero, and hence the shielding efficiency is zero. The air-gap is so located that the radii of the two cylinders are in geometric progression.

Experimental Data

From the discussion under "Theory" it is evident that although the equations (2 to 11 incl.) were derived with the assumption of a steady magnetic field the above calculated values apply equally well to an alternating magnetic field if the effective permeability is used. However, the results are far from sufficient to determine the shielding efficiency of a magnetic shield for a transformer. The shielding due to the eddy currents, i_{e1} (see Fig. 1), is not included. To include this the formulæ (12) to (20) inclusive must be used. The equations have been derived with the assumption that the length of the cylinders is infinite. In practical applications this is obviously not so. Covers, however, approximately counterbalance the effect of the finite length of the cylinders. The magnetic core of the transformer also materially affects the shielding efficiency. In connection with such factors as these which would be very laborious to treat theoretically some experimental information will now be given. The frequency range of the disturbing magnetic field was limited to from 50 to 4000 cycles.

The shielding efficiency has previously been defined as follows:

$$\text{Shielding Efficiency} = 20 \log_{10} H_e/H_i. \tag{1}$$

From the standpoint of the shielding of transformers we are primarily interested in the reduction of the transformer terminal voltage which is caused by the disturbing magnetic field. For this reason it will be found convenient to define the shielding efficiency in connection with transformers in decibels as follows:

$$\text{Shielding Efficiency} = 20 \log_{10} E_e/E_i, \tag{21}$$

where E_e is the terminal voltage due to the disturbing magnetic field with the shield removed and E_i the corresponding voltage with the transformer inside the shield. In addition, E_e and E_i are restricted to the maximum terminal voltages, with respect to position, that is, the transformer is assumed to be in that angular position with respect to the direction of the magnetic field, in which the maximum terminal voltage is obtained. With an unshielded shell type transformer, for example, this position would be that in which the axis of the winding coincides with the direction of the disturbing magnetic field. This restriction is necessary for the definition (21) to be of any value.

With a small air core coil and an infinite cylinder the axis of which is perpendicular to the axis of the winding and to the direction of the disturbing magnetic field the two definitions are equivalent.

The circuit used in making measurements consists of a field coil producing a magnetic field, a pick-up coil and a vacuum tube voltmeter. The field coil is a loop two feet in diameter consisting of 500 turns of wire. The magnetic field at the center of this coil is uniform over a space sufficiently large for the purpose. The pick-up or search coil when placed in the magnetic field will have a voltage produced across its terminals. This voltage is measured with the vacuum tube voltmeter and E_e and E_i in equation (21) are thus obtained. The size of the shell type core on which the pick-up coil is wound is $3'' \times 1\ 29/32'' \times 1''$ where the $1\ 29/32''$ dimension is parallel to the axis of the winding, the $1''$ dimensions being the pile-up of laminations. Unless otherwise mentioned this coil was used in all of the following measurements.

Permalloy having an initial permeability of the order of 5000 at low frequencies was employed for both the core and the shield throughout this investigation.

Permalloy Cases

In Fig. 7 is shown the shielding efficiency vs. frequency of a rectangular permalloy case which consists of five contiguous layers of

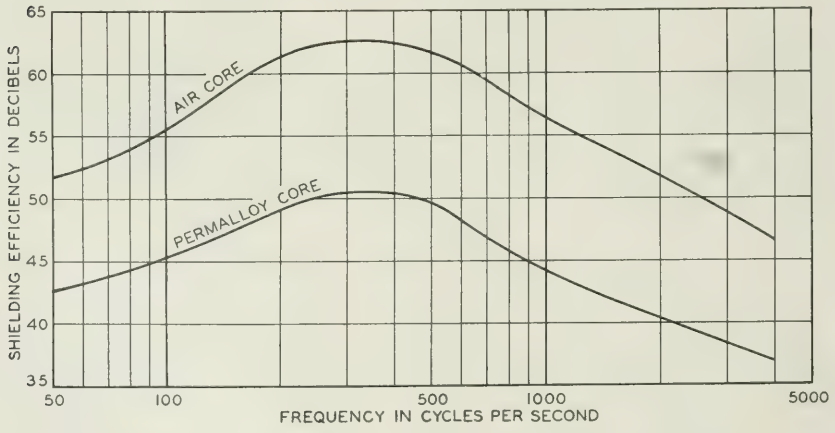


Fig. 7—Observed shielding efficiency of laminated permalloy case.

.014″ thick permalloy sheet. The size of this case is approximately $2\ 3/8'' \times 2\ 1/2'' \times 3\ 1/4''$ and the position of the coil inside the case is such that the axis of the winding is parallel to the $2\ 1/2''$ dimension. The relative size of the coil and case is such that there is approximately

1/8″ clearance between the core of the coil and the case. The shielding efficiency is given for a coil having a permalloy core and also for a coil having a non-magnetic core. These curves illustrate the effect of the magnetic core upon the shielding efficiency. At low frequencies the shielding is mainly due to the magnetic properties of the shield material. As the frequency increases, however, the effect due to the eddy currents, i_{e1} (Fig. 1), increases and the shielding efficiency increases. At approximately 300 cycles a maximum is reached and from here on up to 4000 cycles the shielding efficiency decreases. This is due to the fact that at these frequencies the eddy currents, i_{e2} (see Fig. 1), decrease the effective permeability of the material at a greater rate than the shielding efficiency is increased due to the eddy currents i_{e1}. The slope of the curves at 50 cycles is not zero. This shows that even at 50 cycles there is a considerable shielding effect due to the beneficial eddy currents.

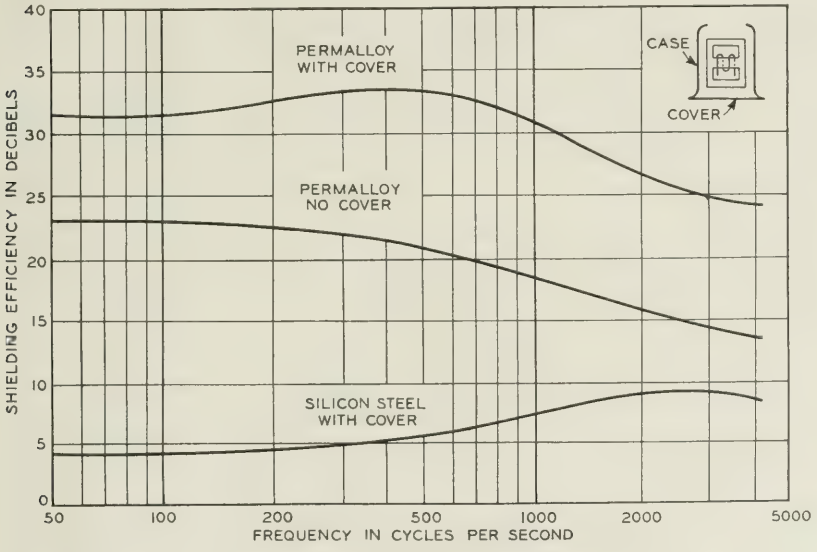


Fig. 8—Observed shielding efficiency of cases made of 1/32″ thick permalloy and silicon steel sheet.

The curves of Fig. 8 show the shielding efficiency vs. frequency of a cylindrical permalloy case, the thickness of the walls of which is 1/32″. The approximate dimensions of this case are 3 1/4″ high × 2 5/8″ diameter and the relative dimensions of the case and core are such that there is approximately 1/8″ clearance between the core and the case. A comparison between the two curves for permalloy shows the effect of the cover upon the shielding efficiency.

A comparison between the curve for permalloy and that for silicon steel on Fig. 8 gives a striking example of the advantage of using permalloy instead of steel.

Effect of Air-Gap Between Core and Shield

It has been pointed out previously that the magnetic core decreases the reluctance through the interior of the case and in that way decreases the shielding efficiency. When the size of the core is such as to almost fill the case, that is, when the air-gap between the core and the case is small, this effect becomes large. In Fig. 9 are given some

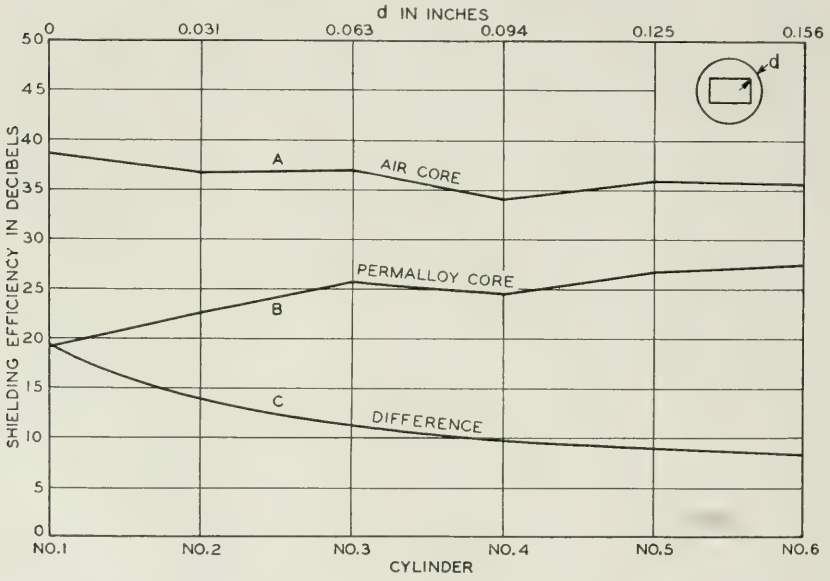


Fig. 9—Observed effect of air-gap between core and shield. Cylindrical shield. Frequency 70 cycles.

data in this connection for cylindrical shields. Six permalloy cylinders were used in this illustration. They are so constructed as to fit over each other, the smallest fitting snugly over the transformer core ("d" = 0). Each cylinder consists of two layers of .014" thick permalloy sheet. The cylinders have been numbered 1 to 6 from the smallest to the largest, respectively. Curve "A" of Fig. 9 shows the shielding efficiency with a non-magnetic core and curve "B" gives the corresponding information with the permalloy core having a permeability of approximately 5000 at 70 cycles and low field strengths which are the conditions under which the measurements were made.

These curves have been drawn discontinuously because the shielding efficiency is a function not only of "d" but of other factors such as permeability, size of the cylinders, etc. Curve "C" on the other hand is primarily a function of "d" and has, therefore, been drawn continuously. This curve gives the difference between the shielding efficiency with an air core and with a permalloy core and shows the advantage of increasing the air-gap between the core and the cylinder. Thus, for example, in this particular case approximately 8 db better shielding efficiency is obtained with an air-gap of 1/16" than if there is no air-gap.

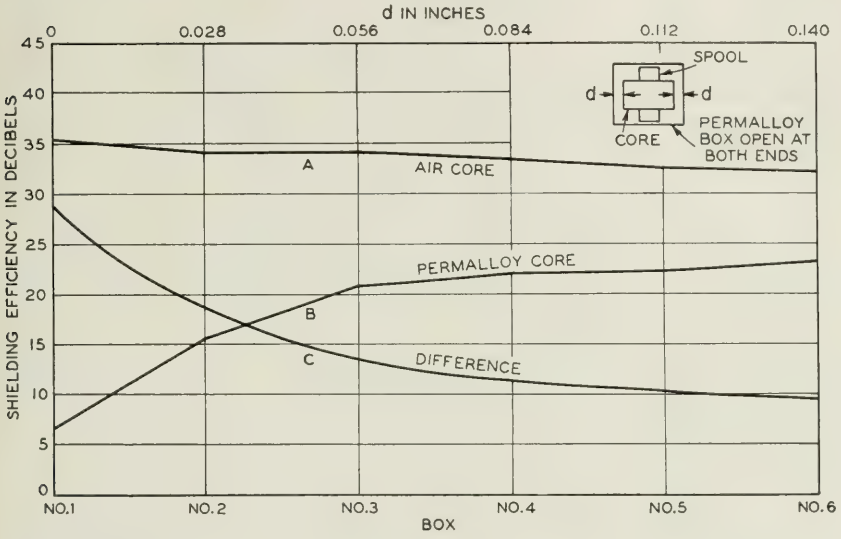


Fig. 10—Observed effect of air-gap between core and shield. Rectangular shield. Frequency 70 cycles.

Similar curves are given in Fig. 10 for rectangular boxes of the same height and same wall thickness as the cylinders. The measurements were made at 70 cycles per second. Due to the larger contact area between the core and the shield the effect of the air-gap is much greater here than with cylinders.

As the effective permeability of the walls of the shield increases the effect of an air-gap increases, other things being equal. In general it may also be said that as the shielding efficiency increases (due to increased thickness of the walls, for example) the effect of an air-gap increases.

If there is an appreciable air-gap between the core and the shield a large variation in the effective permeability of the core will affect

the shielding efficiency very little. That is, practically the same results will be obtained with a silicon steel core, having a permeability of 400 as with a permalloy core having a permeability of 5000. On the other hand if the silicon steel core is replaced by an air core the change in the shielding efficiency will be of such an order as is indicated by Fig. 7. The reason for the small effect of changing from a permalloy core to a silicon steel core as compared to the changing from a silicon steel core to an air core is, of course, due to the fact that in the first case there is a decrease in permeability of 12.5 : 1 while in the second case the corresponding reduction in permeability is 400 : 1.

High Efficiency Shields

A shielding efficiency of from 20 to 50 db is, for many purposes, sufficient in connection with the shielding of transformers. Occasions arise, however, when a shielding efficiency much greater is desired.

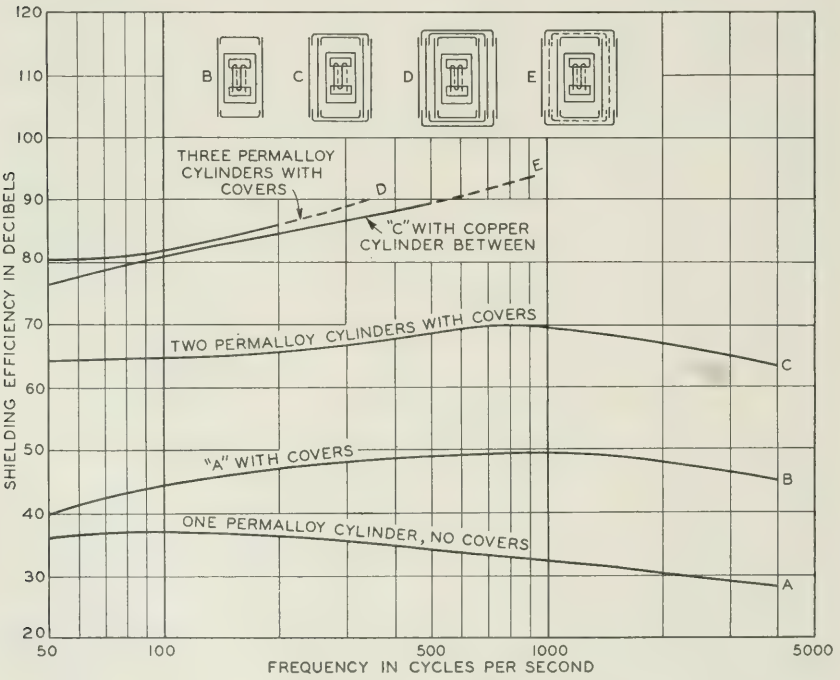


Fig. 11—Observed shielding efficiency of various high-efficiency shields.

To accomplish this by means of a simple case, magnetic material having a permeability much greater than is now readily available would be needed. The curve *B* of Fig. 11 gives the shielding efficiency

of a permalloy cylinder which has a permeability of approximately 5000. This cylinder is 4.5″ high, inside diameter 2.5″, and thickness of wall equal to .07″. By increasing the thickness of the wall the efficiency would be only slightly increased. This is evident from a study of Fig. 4. However, by placing a second cylinder over "*B*" a substantial improvement is obtained (Curve *C*). Still greater shielding efficiency is obtained by placing a third cylinder over the former two as shown by curve *D*. The dimensions of the second and third cylinders are such that the ratios between the outside and inside radii of the three cylinders and of the air-gaps between them are approximately in geometric progression. The height of the second and third cylinders is also 4.5″ and the effective permeability at low frequencies and field strengths approximately 5000.

Since the effective permeability of the permalloy used in the above shields is close to 5000 we can compare these data with the theoretical curves of Fig. 4, which were calculated with the permeability assumed equal to 5000. This comparison shows that the theoretical analysis of the shielding of infinite cylinders against steady magnetic fields may be applied to the shielding of transformers. Due to such factors as a magnetic core inside the shield, eddy current shielding, end effects etc. only an approximate check can be expected. At 50 cycles per second the measured values for one cylinder and for two and three concentric cylinders are 40, 64, and 80 db respectively. Corresponding calculated values as given by Fig. 4 are 41.5, 66, and 89 db respectively.

It is evident from the effect of the copper cylinder between two permalloy cylinders as shown by curve *E* (Fig. 11) that three permalloy cylinders with copper cylinders between the inner and middle and between the middle and outer will give a shielding efficiency of the order of 100 db (voltage ratio $E_e/E_i = 10^5$) from 50 to 4000 cycles.

Effect of Covers

The information given in Fig. 12 shows the importance of covers. This figure gives the shielding efficiency of a cylinder which consists of two layers of .014″ permalloy sheet. Curve *A* gives the shielding efficiency without any covers and curve *B* shows the advantage of adding covers which overlap 1/2″ and consist of two layers of .014″ permalloy sheet. The two curves *C* give the shielding efficiency of the same cylinder provided with flat plate covers, C_1 representing covers .014″ thick and C_2 covers .028″ thick. The relative size of the coil and cylinder is such that there is a clearance of approximately 1/16″ between the core and the magnetic shield. It is evident that in this particular instance the covers are very important. Regarding

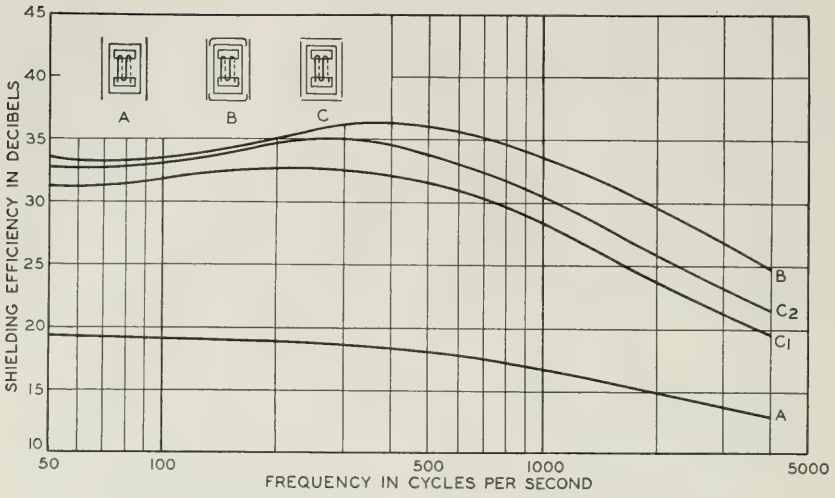


Fig. 12—Observed effect of covers on cylindrical shields.

contact between the cover and the cylinder it may be shown that at low frequencies this is immaterial while at higher frequencies the reverse is true. However, even at these higher frequencies a small overlap (as for *B*) is sufficient.

Examples of the effect of covers upon the shielding efficiency are also given by Figs. 8 and 11. The effect of covers on a copper cylinder as shown by curves *A* and *B* (Fig. 13) is of special interest.

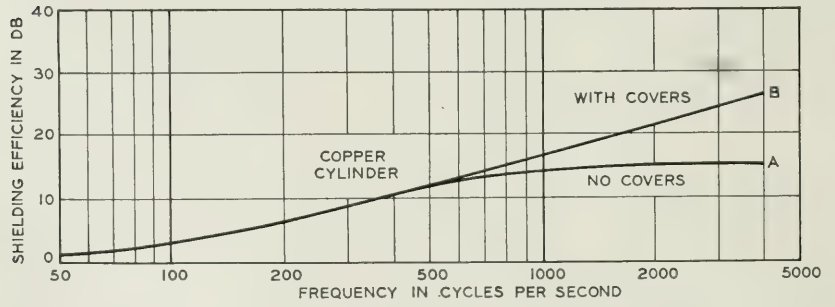


Fig. 13—Observed shielding efficiency of a copper cylinder.

Use of Copper

The curves *A* and *B* of Fig. 13 give the shielding efficiency vs. frequency of a cylinder made of copper. This cylinder has an inside diameter of 2 5/8″ and is 4.5″ high. The thickness of the wall is 1/16″. The shielding efficiency is here entirely due to the eddy

currents, i_{e1} (Fig. 1). Curve B shows that, after the effect due to the open ends of the cylinder is eliminated by means of covers, the shielding efficiency is approximately proportional to the logarithm of the frequency.

Although copper has a very low shielding efficiency at low frequencies when used alone, tests show that under certain conditions it is very effective when used in conjunction with permalloy. This is illustrated by a comparison between the curves C and E of Fig. 11. The copper cylinder is similar to the one referred to above except that the thickness of the wall is only 1/32″. The permalloy cylinders are those for which the shielding efficiency is given by curve C of the same figure.

Another striking example of the use of copper in conjunction with permalloy is furnished by Fig. 14. The curve A gives the observed

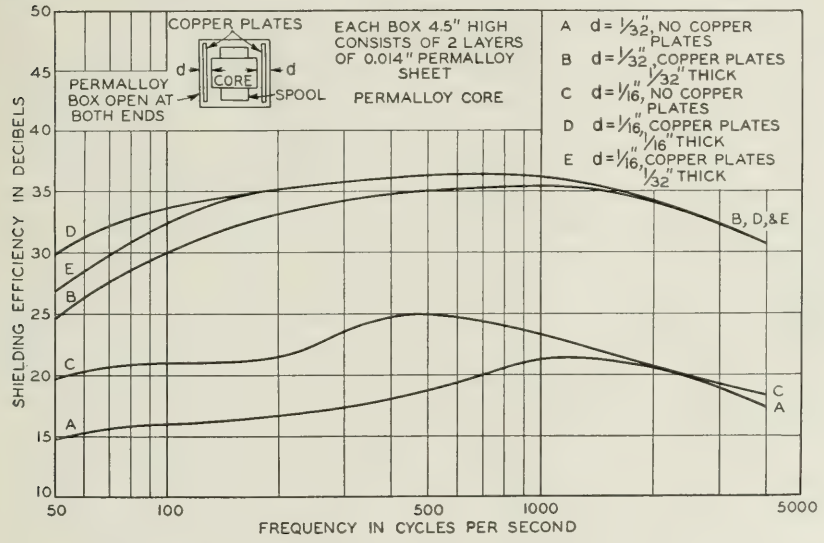


Fig. 14—Observed effect of copper between core and permalloy shield.

shielding efficiency of a permalloy box when there is an airspace between the core and the box of 1/32″ (d = 1/32″). If 1/32″ thick copper plates are inserted between the core and the box (see Fig. 14), there is a great improvement in the shielding efficiency as a comparison between the curves A and B shows. This improvement is 10 db at 50 cycles although the effect of the copper plates alone would be of the order of one db, as is evident from the curves on Fig. 13. At higher frequencies the improvement is still better. It is approximately 15 db between 100 and 4000 cycles. Approximately the same results

are obtained with a spacing of 1/16″. A comparison between the curves *C* and *E* shows the improvement which is obtained in this case with 1/32″ copper plates. The effect is somewhat less than with a spacing of 1/32″, at least at frequencies below 1000 cycles. At low frequencies copper plates, 1/16″ thick, show a slight improvement over the 1/32″ copper plates as is shown by the curve *D*.

Curve *B* on Fig. 10 shows that if the airspace between the core and the box is small the shielding efficiency is very low. In a case like this a copper spacer is very effective. For example, a 5 or 10-mil copper plate replacing an airspace of the same thickness will greatly improve the shielding efficiency.

General

The foregoing is a discussion of the magnetic shielding of transformers from external magnetic fields. The reverse problem of shielding a transformer or coil so as to prevent its magnetic field from affecting other apparatus has not been considered. However, it is safe to assume that approximately the same degree of shielding will be obtained, provided the leakage field does not produce excessive saturation in the shield. That is, assuming that a power transformer is producing a disturbing magnetic field in the space occupied by an input transformer, then a shield over the power transformer will produce approximately the same effect as a shield over the input transformer, where each shield has been constructed in accordance with the information on the foregoing pages. This has been demonstrated experimentally in an article by J. E. R. Constable in the *Wireless World* of February 26, 1937.

Although this paper has been restricted to the magnetic shielding of transformers it is equally applicable to any apparatus which is susceptible to inductive pick-up. This is because in any apparatus where there is inductive pick-up there is in effect a coil. It may be an actual coil and it may be only a loop of lead wires.

The author wishes to thank Mr. E. T. Hoch for many helpful suggestions.

Bibliography

1. J. Stefan, *Wied. Annalen*, Vol. 17, p. 928, 1882.
2. A. W. Rücker, *Phil. Mag.*, Vol. 37, p. 95, 1894.
3. H. DuBois, *Wied. Ann.*, Vol. 63, p. 348, 1897; Vol. 65, p. 1, 1898, also *Electrician*, Vol. 40, 1897, pp. 218, 316, 511, etc. cont.
4. A. P. Wills, *Phys. Rev.*, Vol. 9, p. 193, 1899; Vol. 24, p. 243, 1907.
5. James Russell, *Roy. Soc., Edinb., Trans.*, Vol. 40, p. 631, 1903.
6. E. F. Nichols and S. R. Williams, *Phys. Rev.*, Vol. 27, p. 250, 1908.
7. W. Esmarch, *Ann. der Phys.*, Vol. 39, p. 1540, 1912.
8. W. W. Coblens, *Bul. Bu. of Standards*, Vol. 13, p. 423, 1916.

9. C. Benedicks, *Ann. der Phys.*, Vol. 72, p. 236, 1923.
10. R. H. Barfield, *Inst. Elec. Eng.*, Vol. 62, p. 249, 1924.
11. D. W. Dye, *Jour. Sci. Inst.*, Vol. 3, p. 65, 1925.
12. A. V. Hill, *Jour. Sci. Inst.*, Vol. 3, p. 335, 1925.
13. J. H. Morecroft and A. Turner, *I. R. E., Proc.*, Vol. 13, p. 477, 1925.
14. H. Pleijel, Svenska Ingeniörsvetenskapsakademiens Handlingar, NR 49, 1926.
15. H. L. Curtis, *A. I. E. E. Journal*, Vol. 48, p. 453, 1929.
16. S. L. Gokhale, *A. I. E. E. Trans.*, Vol. 48, p. 1307, 1929.
17. N. Hillers, *Telefunken Zeitung*, Vol. 13, pp. 13–28, 1932.
18. L. V. King, *Phil. Mag.*, Vol. 15, p. 201, February 1933.
19. W. Lyons, *I. R. E., Proc.*, Vol. 21, p. 574, April 1933.
20. G. W. O. Howe, *Wireless Engr. and Exp. Wireless*, Vol. 11, p. 347, July 1934.
21. S. A. Schelkunoff, *Bell Sys. Tech. Jour.*, Vol. 13, p. 532, October 1934.
22. *Radio Engg.*, Vol. 15, p. 11, July 1935.
23. T. E. Sterne, *Rev. Sci. Inst.*, Vol. 6, p. 324, October 1935.
24. W. F. Randall, *The Nickel Bulletin*, Vol. 9, No. 4, p. 73, April 1936.
25. R. Bachstroem, *Arch. f. Elektrotechnik*, Vol. 30, p. 267, April 1936.
26. S. A. Schelkunoff, *Radio World*, Vol. 29, p. 47, April 1936.
27. Samuel Levy, *I. R. E., Proc.*, Vol. 24, p. 923, June 1936.
28. C. W. Oatley, *Phil. Mag.*, Vol. 22, p. 445, September 1936.
29. J. E. R. Constable, *Wireless World*, Vol. 40, p. 198, February 1937.
30. W. Herzog, *E. N. T.*, Vol. 14, p. 81, March 1937.

Coaxial Cable System for Television Transmission*

By M. E. STRIEBY

THE reports which have been made on the progress in television development increase the expectation that the broadcasting of visual programs will soon be realized. In anticipation of that result, the Bell Laboratories has been engaged for some time in the development of wire line circuits for transmitting television signals between studios and broadcasting transmitters, or between cities, as may some day be required if television follows in the footsteps of sound program broadcasting.

The wide frequency bands required for television and the dearth of available frequencies appear to force the broadcasting of television signals into the ultra-high frequency range. At these high frequencies, the coverage which can be obtained from a broadcasting station is very limited as compared to that obtainable in the sound broadcasting frequency range. Hence, if television programs are to reach large sections of the country simultaneously, the provision of interconnections between large numbers of television broadcasting transmitters will become even more important than it is today for sound broadcasting.

Coaxial cables have received much publicity as transmission lines for television. The original conception and use of the coaxial form of cable was first as a low frequency submarine conductor and later as a lead-in for radio antennas. The idea of a coaxial cable or other medium for the transmission of very broad frequency bands originated in the course of telephone development in America.[1] The first lengths of such cable for broad-band transmission were made here and its first use for the transmission of a large number of simultaneous telephone conversations was between New York and Philadelphia.[2] In this country the important reason for developing coaxial cable systems was, and still is, that they appear to offer material economies in the provision of large groups of long distance telephone facilities. Television has been secondary.

Recently experiments have been made on the transmission of television signals over the coaxial cable between New York and Philadelphia. This cable contains two coaxial conductor units within

* Presented at A. I. E. E. Winter Convention, New York City, Jan. 27, 1938. Published in June 1938 issue of *Elec. Engg.*

a lead sheath about $\frac{7}{8}''$ in diameter as indicated in Fig. 1. Two coaxial units were provided because, for long distance telephone operation four-wire operation is preferable, one coaxial being employed for transmission east to west and the other west to east. Each coaxial unit is made up of a 13-gauge inner conductor on which hard rubber disks have been placed at intervals of $\frac{3}{4}$ of an inch. The outer metallic tube is made up of 9 overlapping copper tapes so designed that they form essentially a solid copper tube about 20 mils in thickness and .267″ in inside diameter.

The transmission loss of this circuit as a function of frequency is shown on Fig. 2 together with the portion of the attenuation that is contributed by conductance losses. Inasmuch as the intention is to use these conductors at very high frequencies, a high grade insulating material was used with the result that the conductance losses are small.

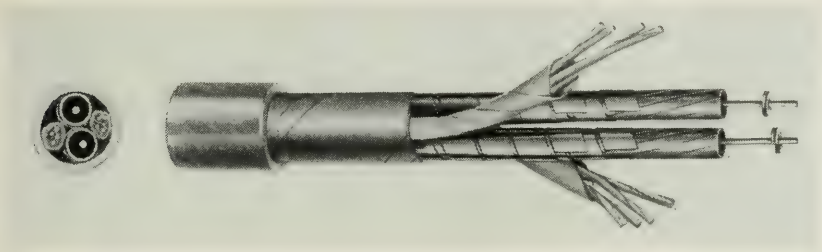


Fig. 1—Section of the New York-Philadelphia coaxial cable.

It should be noted that the attenuation increases very nearly as the square root of the frequency.

In order to transmit high frequencies over long distances, a great deal of amplification is obviously required. The New York-Philadelphia cable was initially equipped to handle a band of one million cycles. Its overall attenuation at a million cycles is approximately 600 decibels. In order to reduce this to a usable amount 10 repeater points were provided at intervals of about 10 miles each having an amplification at the top frequency of about 60 decibels. These repeaters were so designed that they provided less gain at low frequencies than at the high frequencies, in approximately the same degree as the line had less attenuation. To make up for certain cumulative irregularities an equalizer was built and added to the overall circuit. The net result was a transmission path which had approximately zero loss over the whole band which it was desired to use from 60 kc. to 1000 kc. as shown in Fig. 3.

Another complicating factor is, however, involved, as is the case with all long wire circuits, namely, the variation encountered with changes in temperature. The loss in a 10-mile repeater section varies materially from summer to winter. If the cable is hung overhead, this variation is about as shown in Fig. 4 and amounts to a change of about $\pm$ 7 per cent in the attenuation. If the line is buried underground at the normal depth used for telephone cables, the actual variation is about one-third as much. Automatic transmission regulators were developed to compensate for these changes. These

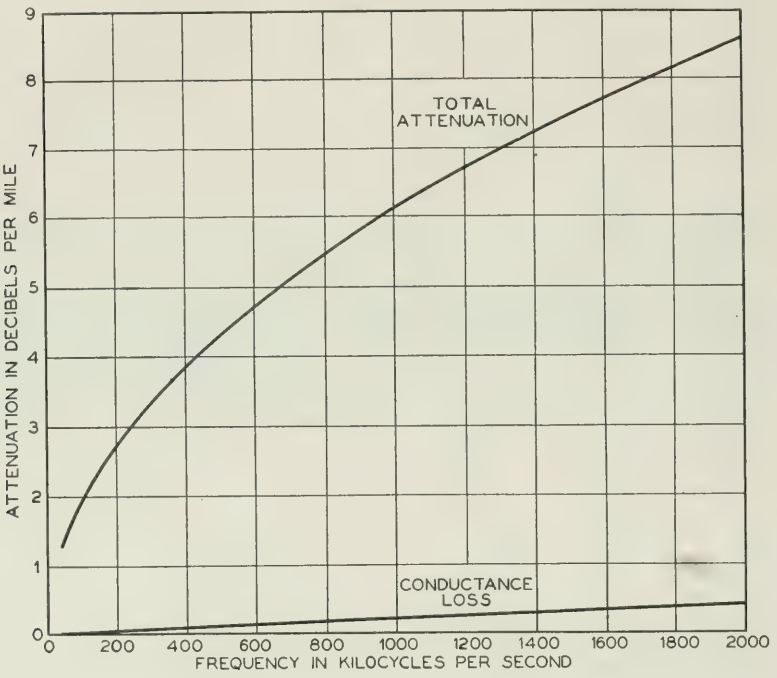


Fig. 2—Attenuation per mile of the New York-Philadelphia coaxial cable. Also the proportion of that attenuation due to conductance losses.

regulators depend, for their operation, on the transmission of a pilot channel. At each point where it is desired to regulate the transmission, the pilot channel is selected by a very narrow band filter and its amplitude used to control an automatic device which changes the gain of the repeater until the amplitude of the pilot at the output of the repeater reaches a certain predetermined value. The regulators on the New York-Philadelphia circuit have operated with such accuracy that it has been unnecessary to make manual adjustments to take care of temperature changes.

The repeaters used along this route were novel in that most of them were placed in small iron boxes located at convenient points along the line. Power for their operation was transmitted at sixty cycles over the coaxial cable itself. Figure 5 shows one such unattended repeater located near Dunellen, New Jersey.

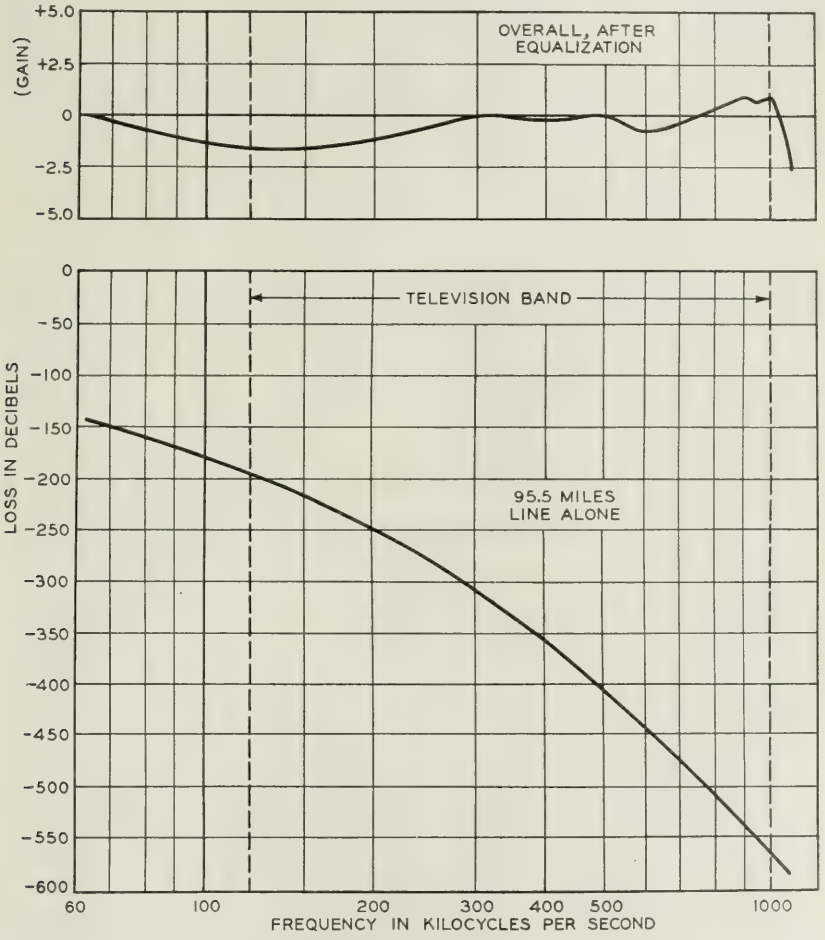


Fig. 3—Overall attenuation of the New York-Philadelphia cable without repeaters and the net attenuation after repeaters and equalizers had been provided.

The coaxial system between New York and Philadelphia was designed to provide 240 telephone circuits. Skeleton terminal apparatus was installed at New York and Philadelphia to test out such operation. This apparatus has been described [2, 3, 4] in various papers and will not be discussed here. Suffice it to say that methods and

equipment have been developed which enable a wide band to be split up so as to obtain hundreds of telephone channels.

For television transmission a quite different problem existed—namely, can very wide band systems be used for the long distance transmission of these complicated signals? In planning tests to be significant of the operation of the cable system for such signals, it was important to obtain as nearly as possible an ideal television signal and as nearly as possible an ideal television receiver. In this

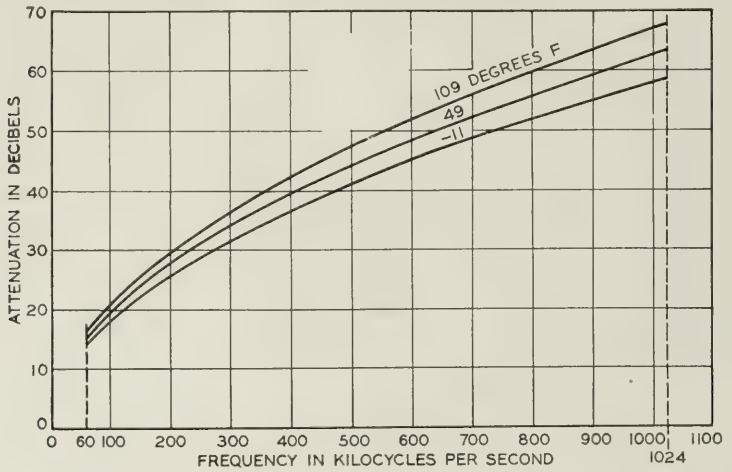


Fig. 4—Attenuation of the New York-Philadelphia type of cable under widely different temperature conditions.

way it was hoped that any defects in the cable transmission itself could be discovered.

Signal Generator

Although television implies the transmission of an actual scene it is much more satisfactory for engineering studies to transmit a motion picture, since exactly the same picture can then be transmitted over and over again as the circuit elements are changed or adjusted. Moreover, it was decided to use mechanical scanning to obtain the most nearly perfect signal possible, and with this form of scanning a film can be much more brightly illuminated than an actual scene and hence is much easier to use. Because of these factors a motion picture film was chosen as the material for the recent experiments.

The scanning disk used in these tests was developed under the direction of Dr. H. E. Ives at the Bell Telephone Laboratories. It consists of a six-foot disk with a circle of 240 lenses near its outer edge.

The arrangement is indicated schematically in Fig. 6. Light from a powerful incandescent lamp behind the disk, passing through one lens at a time, is focussed by the lens to form on the film a small dot of

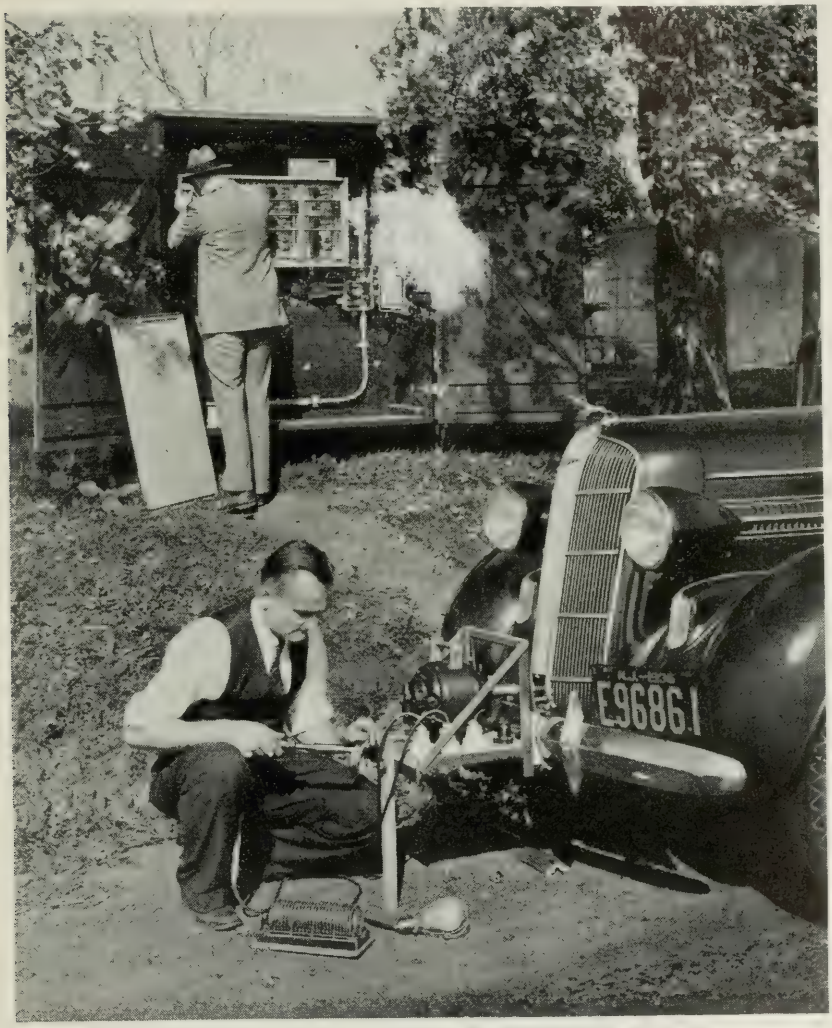


Fig. 5—Repeater near Dunellen, New Jersey.

light about three thousandths of an inch square. The lenses in the disk are spaced by a distance equal to the width of the picture, or a little less than an inch, and as the disk rotates, each spot is moved rapidly across the picture. The film is carried at a uniform rate

downward behind the disk at such a speed that the successive holes throw their light in successive rows across the picture one above another. The film moves one frame for each revolution of the disk. A photosensitive surface mounted behind the film picks up the light transmitted through it, and produces a complex electric current corresponding to the variations of light in the picture. Figure 7 is a photograph of the housing in which the disk is mounted. This

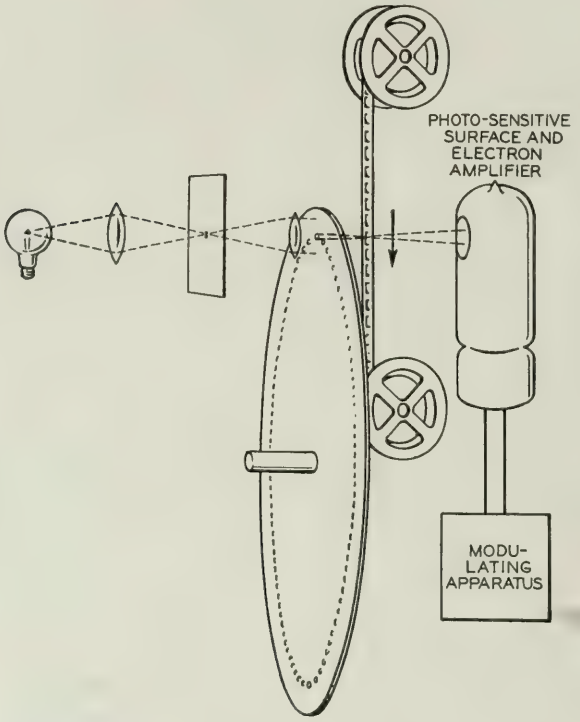


Fig. 6—Schematic diagram of the mechanical scanning arrangement used for television testing.

scanning arrangement produced a picture of 240 lines, 24 frames per second. It was recognized that 24 frames per second were not sufficient to avoid flicker but this choice simplified the scanning apparatus and it was believed would not interfere with engineering tests.

Signal Frequency Range

In order to understand what frequency is required to transmit an image scanned in this way consider the diagram shown in Fig. 8.

Of course, actual television will not ordinarily deal with such a picture, but by means of it an approximate visualization of the problem can be obtained. A desired definition of 240 lines was chosen and it was decided that the requirement would be to transmit square picture elements as indicated in the figure. The shape of the picture which

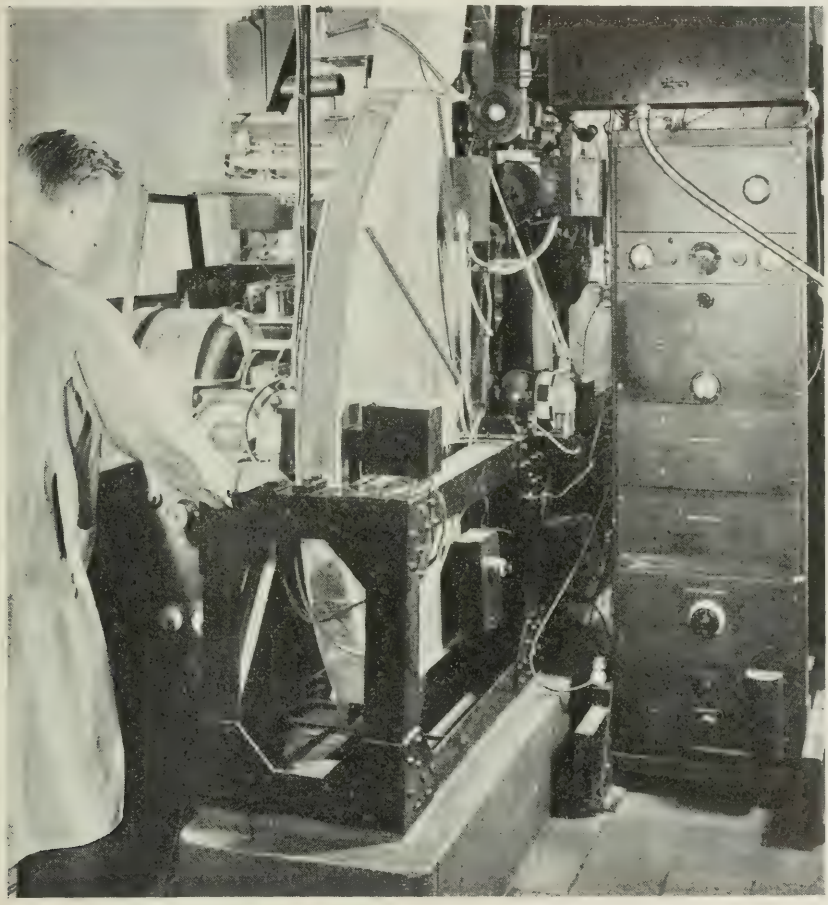


Fig. 7—Photograph of the mechanical scanning disk and associated experimental apparatus.

it was convenient to use with this scanning disk is wider than it is high in the ratio of 7 : 6. This differs somewhat from the standard aspect ratio of 4 : 3, but is easily taken into account. The total number of picture elements in a frame is then $240^2 \times \frac{7}{6} = 67200$.

If the smallest picture element to be transmitted is a single block then the distribution of light and shade over the block is unimportant. The average brightness over the block is what counts. Obviously, a simple approximation is a sine wave as shown at the bottom of the picture. This wave has ½ cycle for each block and is as high a frequency as there is any profit in transmitting for this diagram.

The top frequency needed for such a picture can then be calculated. The number of square elements in the picture computed above is

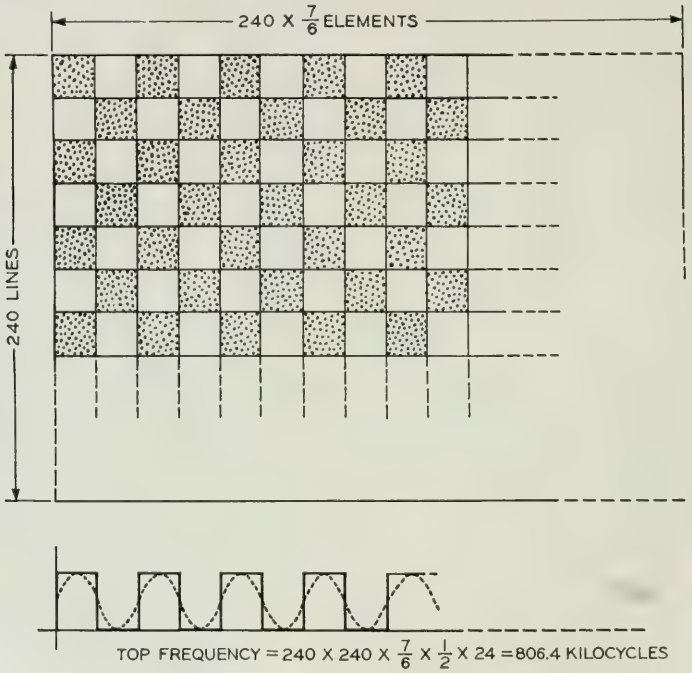


Fig. 8—Diagram illustrating the resolution of an image into picture elements and the derivation of maximum frequency required for transmission.

67,200. As each of these elements represents ½ cycle, this figure is divided by 2, giving 33,600 cycles per frame. As similar frames are reproduced 24 times a second, the result is 24 × 33,600 or 806,400 cycles per second. In a real moving picture, other frequency components may exist at all other frequencies from 800 kc. down to and including direct current. The direct current or zero frequency component controls the general level of brightness of the picture. Where the general level of brightness changes slowly, it results in a component of very low frequency. A composite picture can be

imagined which will produce a pronounced component at any given frequency, hence, it is deemed important to transmit the entire band from 0 to 806 kc.

Receiving Device

At the receiving end an effort was made to obtain as high a degree of fidelity of reproduction as possible. No small factor in the success of the recent experiment was the special cathode ray tube, designed by Dr. C. J. Davisson and used at the receiving end to display the transmitted picture. Some of the features of this tube are indicated schematically in Fig. 9. A stream of electrons from the cathode passes through a series of electron lenses which focus a narrow beam on an aperture .006″ square. Between the lenses and the aperture, however, are two modulating plates connected to the incoming circuit in such a way that there appear on these plates potentials varying according to the voltage of the incoming signals. The effect of

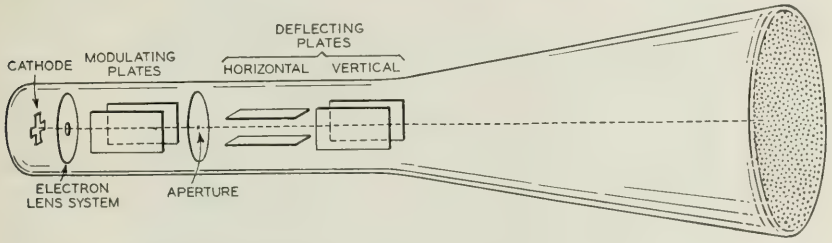


Fig. 9—Schematic diagram of the special cathode ray tube used for viewing transmitted images.

potential on these plates is to deflect the electron beam, and the conditions are such that at maximum strength of signal practically the entire stream of electrons passes through the hole and forms a brilliant spot of light on the front of the tube. As the signal decreases in strength, the electron stream is more and more deflected; less electrons pass through the aperture, and the illumination on the sensitized end of the tube decreases.

In addition to these modulating plates, and placed between the aperture and the front of the tube, are two other pairs of plates mounted in planes at right angles to each other. The potential on one of these sets of plates, controlled by a frequency of 5760 cycles, which is the frequency at which successive lines are scanned, varies in such a way that the beam of electrons passing through the aperture is swept across the front of the tube from left to right, exactly in synchronism with the scanning beam at the sending end. After the beam reaches the farther side of the picture, the potential on the

plates is suddenly changed, and the beam is rapidly moved back to begin the next line. Due to a black mask down the far side of the film being scanned, there is no signal during this very short period while the voltage on the plates is changed, and thus the electron beam is deflected from the aperture and is not visible on the front of the tube during its return.

The potential on the other pair of plates is controlled at a frequency of 24 cycles per second, which is the rate of scanning successive frames. The effect of the potential on these plates is to deflect the electron beam downward in synchronism with the motion of the film at the sending end. This results in the passage of the electron beam across the front of the tube in successive rows, one below another. After the last row has been scanned, the voltage on the plates is changed and returns to the value that causes the beam to appear at the top line of the tube. A properly synchronized blanking-out pulse is introduced between successive frames of the film, so that no signal is received during this interval, and thus the passage of the electron beam from the bottom to the top of the frame is not visible.

Figure 10 is a photograph of one of these cathode ray tubes. Due to its superior design, the image is very sharp over the entire field and a wide range of brightness is secured. The chief factors in its success are the sharp focusing by the electron lenses, the linear deflection of the beam at the aperture, and the great length of the tube, which makes it necessary to deflect the electron beam over only a narrow angle to cover the 7×8 inch field. Since this trial was a test to determine the capabilities of the system, such matters as size and cost, which would be important with commercial receivers, were not controlling.

Modulation System

The frequency band which was generated at the sending end as noted above was 0 to 806 kc. The coaxial cable system used could not transmit this band, because repeaters were not designed to pass frequencies below about 60 kc. This limitation was incorporated in the original design because the cable offers insufficient shielding to various disturbances at low frequencies. For television transmission it was necessary, therefore, to raise the television signal band to a higher frequency position before attempting transmission over the line. A number of considerations led to the decision to raise the entire frequency band 144 kc. for transmission over the coaxial cable.

Where such a wide frequency band is to be raised by an amount less than the width of the band itself, a single modulation is not generally satisfactory. The products of modulation include the

original frequency band as well as the upper and lower sidebands, so that there will be a confusing jumble of frequencies in the modulator output. For this reason a double modulation method was used for the recent experiments.

Fig. 10—Photograph of one of the special cathode ray receiving tubes.

The modulating scheme employed can be followed with the help of Fig. 11, which shows the two modulating steps at the sending end and the two demodulating steps at the receiving end in four lines beginning at the top. A carrier of 2376 kc. is used for the first modulation, which results in a lower sideband from 1570 to 2376 kc. and an upper

sideband from 2376 to 3182. The carrier itself is eliminated in the balanced modulator. The output of this modulation is passed through a filter, but because the two sidebands touch each other at 2376 kc., the filter cannot cut off all the upper sideband. At the output of this filter there is thus the lower sideband plus a small amount of the lower part of the upper sideband. The upper sidebands from all subsequent modulations are readily eliminated by filters because of the wide separation.

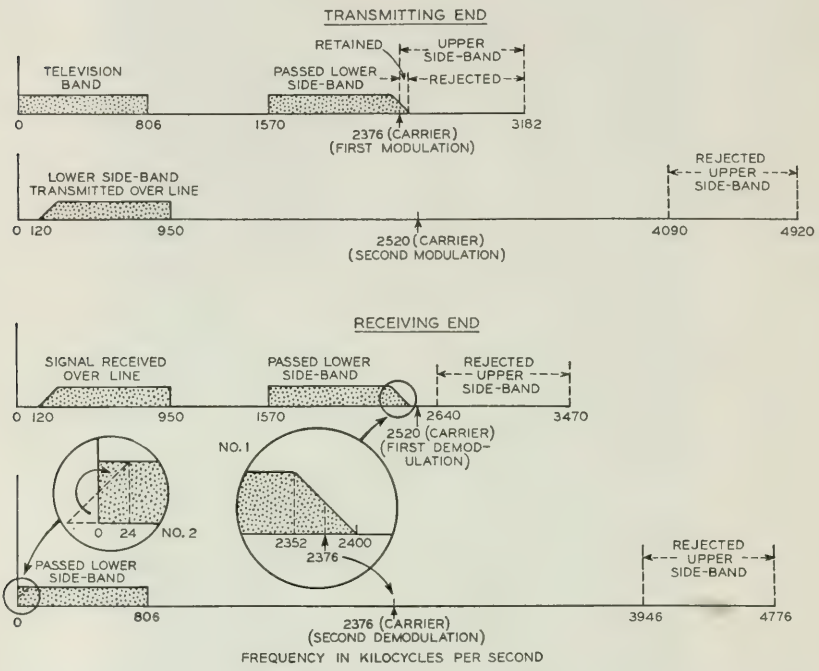


Fig. 11—Schematic diagram illustrating the processes of modulation and demodulation used in transmitting television signals over the coaxial line.

The carrier for the second modulation is 2520 kc., and the lower sideband extends from 950 down to 144 kc. plus a vestigial upper sideband remaining from the first modulation which extends down to 120 kc. The high-pass filter following this modulation is accurately designed to pass with controlled attenuation a group of frequencies just above 144 kc. and the vestigial sideband. The resulting single sideband, extending from 120 to 950 kc., is then passed over the coaxial cable.

At Philadelphia the received signal, together with a carrier of 2520 kc., is applied to the first demodulator, and the lower sideband,

from 2400 down to 1570, is passed to the second demodulator where a carrier of 2376 kc. is applied. The lowest frequency of the lower sideband, 1570 kc., is converted to 806 kc., becoming the highest frequency of the final demodulated band. The frequencies from 2352 to 2400 kc. of the sideband before the second demodulation have been attenuated somewhat by the high-pass filter following the second modulator, and the second demodulating carrier, 2376 kc., falls in the middle of this attenuated band as shown in inset No. 1. Frequencies extending about 24 kc. above the carrier are inverted by the demodulation, and superimposed upon the corresponding frequencies just below the carrier. The magnitude and phase of these components are proportioned by the high-pass filter and an equalizer so that the overall result, when they are superimposed, is an essentially flat transmission band from 0 to 806 kc.

The above steps of modulation involved a number of difficulties. In the first place the signal level must be carefully controlled so that on the one hand it does not sink into the background noise, while on the other hand it must not be raised to such high levels that unwanted modulation products are produced in too great magnitude. The first modulator presents some special problems. It must accommodate all frequencies from 0 to 806 kc. In order to eliminate the carrier, it must be balanced to a very high degree—about 80 db in this case. The reason the carrier must be so completely wiped out is that the 0 frequency component of the signal is identical with the carrier at the output and hence the true d-c. value of the signal must be exceptionally free from carrier interference.

Referring to Fig. 12, the next piece of apparatus is a band filter to eliminate the video signal and cut off the top edge of the band. Then follows the 2nd modulator which is quite conventional. A low-pass filter is next and is very important as it performs part of the function of cutting off and adjusting the vestigial sideband. Then follows an amplifier, a predistorting network to partially equalize the amplitudes of the different components of the signal, an aperture equalizer to correct for the fact that the scanning spot is of finite size, a terminal equalizer to make up for irregularities in the overall setup and other amplifiers.

The carrier apparatus at the sending end is shown on Fig. 13. It is mounted in rather conventional form except for the 1st modulator which was arranged to minimize the effect of low-frequency vibrations. At the receiving end about the same apparatus is required in the inverse order and will not be discussed in detail.

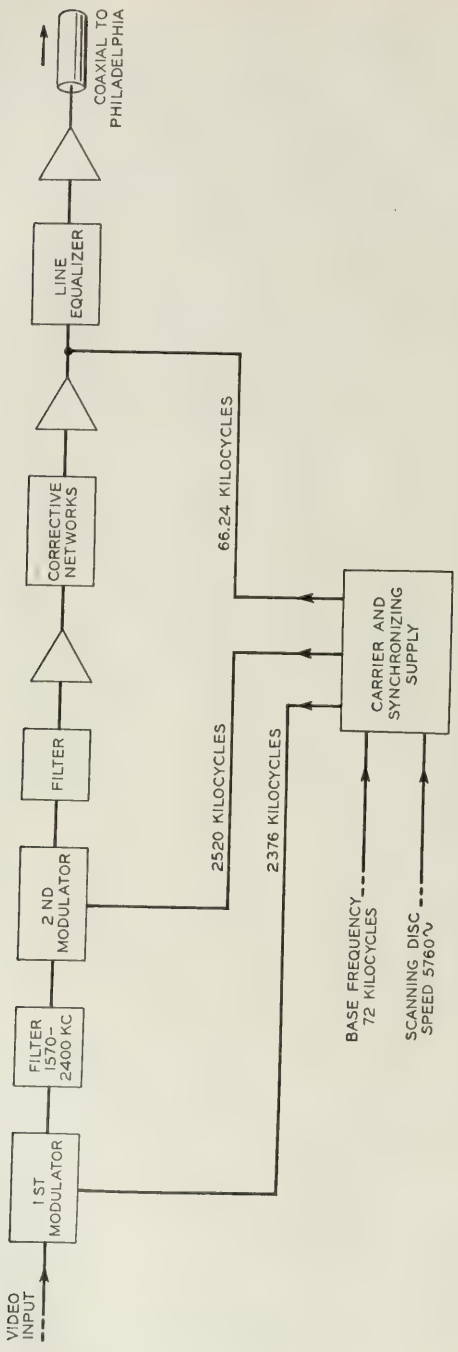


Fig. 12—Block diagram of the transmitting carrier television terminal at New York.

Carrier Supply

Another important feature of this system was the provision of accurately spaced carriers for the various modulating operations. The Bell System 4 kc. standard was used to produce a 72 kc. (18th harmonic) base frequency signal which could be transmitted over the line

Fig. 13—Photograph of the transmitting carrier television terminal with associated sound apparatus.

and so tie the transmitting and receiving ends together. From this, by harmonic generation, the carriers used in the two steps of modulation were produced, being the 33rd and 35th harmonics.

To synchronize the scanning at the receiving end with the transmitting disk required another direct tie. The disk is driven by a d-c.

motor and its speed cannot be kept very constant. By means of an auxiliary lamp and photocell, the frequency with which one scanning line followed another—approximately 5760 per second—was obtained. This was modulated with the 72 kc. mentioned above and transmitted over the line as a lower sideband at 66.24 kc. At the receiving end by demodulation, the exact line speed is obtained and used to drive the horizontal sweep. The vertical sweep is obtained by generating the 240th subharmonic of this—namely 24 c.p.s.

Line Equalization and Test Results

Returning now to the line transmission problem, the signals which might be transmitted in the general case are indicated in Fig. 14. They include the pilot channels used for automatic transmission regulation at 60 kc. and 1024 kc. For convenience, one telephone channel with a carrier at 64 kc. is indicated as an order wire, and a

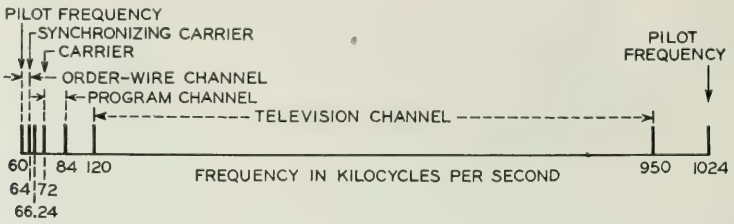


Fig. 14—Frequency allocation for a television transmission system with associated control circuits.

wide-band program channel with carrier at 84 kc. to transmit the sound. These, of course, could be provided with ordinary telephone facilities. The base frequency of 72 kc. and the disk synchronizing sideband at 66.24 are also included. For the television signal itself the band from 120 to 950 kc. is provided. Actually in the tests to Philadelphia, automatic regulation was not needed and a separate wire line was used for synchronization.

It was necessary to provide networks and equalizers to insure that the coaxial line did not distort the ultimate image due to unequal attenuation, resulting in amplitude distortion, or to unequal time of transmission, causing phase distortion. The actual attenuation characteristics of the line [5] and the overall result were shown above in Fig. 3. The requirements for phase distortion are rather difficult to meet. The details in the scanned picture result in various frequencies of the electrical signal, and if these details are to appear in the reproduced picture in the same relative position as in the scanned picture, it is essential that all frequencies be received in very closely the same

relative time relationship as they are generated. Referring back to Fig. 8, it was assumed that no picture element could be displaced by more than about half its width. This led to the decision to hold frequencies between 806,000 and 3000 cycles to a delay distortion of about 0.3 microsecond. For a similar degradation of detail in the vertical direction, the permissible delay distortion is 280 times as great which, in a system of this type, is very easily obtained. The actual circuit roughly met these requirements as indicated by Fig. 15,

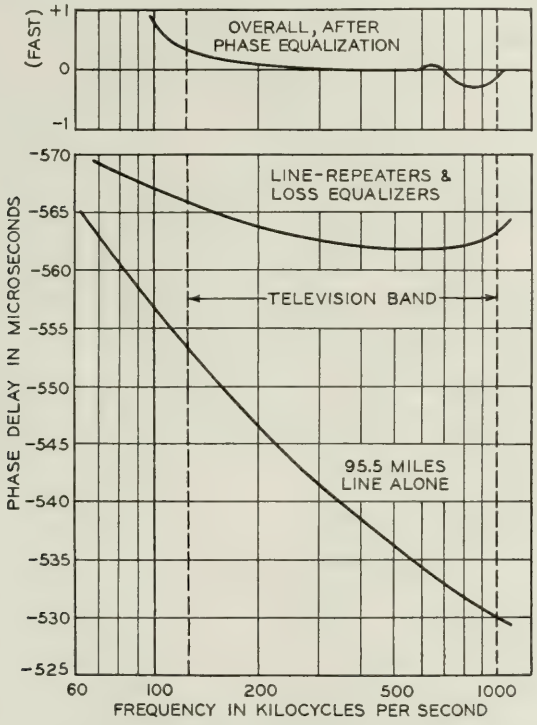


Fig. 15—Phase delay of New York-Philadelphia television circuit.

which shows the phase delay characteristics of the line, repeaters and equalizers, and of the overall circuit at the frequencies used for transmission.

Noise or interference is very annoying in television transmission; and pattern, or single frequency interference, is particularly objectionable. The permissible noise or interference depends on the amplitude range of the reproduced picture. During these experiments, it was found that a substantially linear response could be obtained over a signal current range of about 20 db—a brightness ratio of 10 to 1.

The actual reproduced pictures considerably exceeded this range; in fact a brightness ratio of 50 or 100 to 1 was realized. In these tests it was found desirable to hold random interference down about 40 db below the maximum signal, and pattern interference down at least 15 db more.

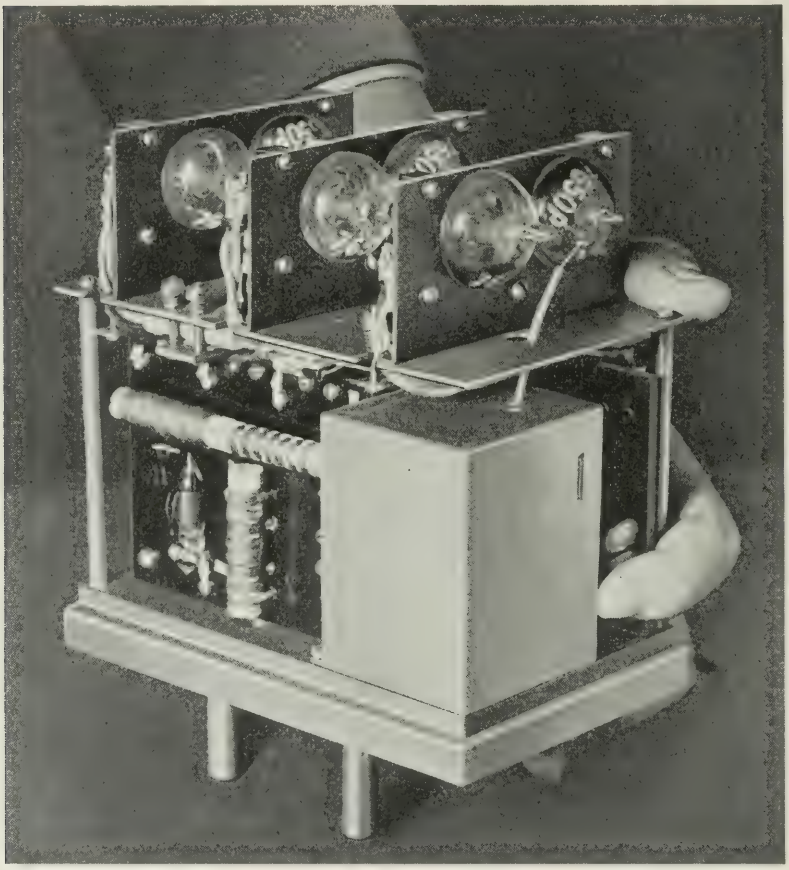


Fig. 16—Photograph of a two million cycle amplifier under development for experiments on the coaxial cable.

The engineers who worked on the system, and outside experts who observed it, expressed the opinion that the reproduced pictures in Philadelphia were substantially the same as those seen on a similar receiving device in New York, thus showing that the cable system itself introduced no appreciable distortion.

Conclusion

As a result of the experimental transmission of the pictures over the coaxial cable from New York to Philadelphia it has been proved that wide-band signals of the type required for television can be satisfactorily transmitted over a coaxial cable system, and that in such transmission the distortion introduced by the wire line circuits can be made so small as to be inappreciable, in its effect on the received picture.

The work on these very wide-band systems has only begun and repeaters and terminal apparatus are now under development capable of handling wider bands of frequency. At the present time work is under way on a two-million cycle system for telephone transmission and a trial installation is being made between New York and Princeton. The system will transmit a frequency band of about two million cycles corresponding to a capacity of 480 telephone circuits. Repeaters on this system will be about 5 miles apart and will consist of unattended boxes somewhat smaller than the one-million cycle repeaters illustrated above and placed either in manholes or on poles along the route. Within these boxes there are placed two amplifiers, one for eastbound transmission, the other for westbound, together with the necessary filters and power supply apparatus. The actual amplifiers themselves are quite small compact units one of which is shown in Fig. 16. Two megacycles, of course, is not a sufficiently wide band to transmit the present R.M.A. standard 441-line television signal, but is a logical step toward more economical telephone circuits. Development is also under way on amplifiers capable of transmitting three megacycle bands of frequency, which should amply satisfy the requirements for transmitting the 441-line television signals now envisioned as standard by the television industry.

References

1. "Systems for Wide Band Transmission Over Coaxial Lines" by L. Espenschied and M. E. Strieby, *Electrical Engineering*, Vol. 53, pp. 1371–1380, October 1934.
2. "A Million-Cycle Telephone System" by M. E. Strieby, *Electrical Engineering*, Vol. 56, pp. 4–7, January 1937.
3. "A Carrier Telephone System for Toll Cables" by C. W. Green and E. I. Green, *Bell Sys. Tech. Jour.*, Vol. XVII, pp. 80–105, January 1938.
4. "Cable Carrier Telephone Terminals" by R. W. Chesnut, L. M. Ilgenfritz and A. Kenner, *Bell Sys. Tech. Jour.*, Vol. XVII, pp. 106–124, January 1938.
5. "Transmission Characteristics of the Coaxial Structure" by J. F. Wentz, *Bell Laboratories Record*, Vol. XVI, pp. 196–200, February 1938.

Stabilized Feedback Oscillators

By G. H. STEVENSON

The author presents a mathematical consideration of the conditions which insure constant frequency of the vacuum tube oscillator under changes of electrode potentials or of the cathode temperature. It has already been shown that the grid and plate resistances may enter into the determination of the frequency. The problem is treated here in the manner suggested in the recent studies of feedback amplifiers. The conditions necessary for stability are developed in terms which are independent of particular circuit configurations and are applicable to certain dissipative circuits as well as to purely reactive systems.

THE frequency deviations that accompany changes of the electrode potentials or of the cathode temperature in many types of vacuum tube oscillators have been recognized for some time as having their origin in the variation of the internal resistances of the tube. Llewellyn has shown [1] that both the grid and plate resistances may enter into the determination of the frequency and, by treating the problem as one of network design, has demonstrated the possibility of making the frequency substantially independent of the tube resistances. He also devised a large number of oscillator circuits stabilized in this way and established the conditions necessary for stabilization in each case.

The problem is treated here in a somewhat more general manner suggested by recent studies of feedback amplifiers.[2] The conditions necessary for stability are developed in terms which are independent of particular circuit configurations and which permit their application to certain types of dissipative circuits as well as to purely reactive systems. While no new fundamental principles are presented, it is thought that the restatement of the known principles in broader terms may be of interest.

The mathematical theory will be developed for the case of the single-tube oscillator circuit, since this is the form most generally used. The extension of the theory to multiple stage circuits presents little or no difficulty. The principal assumptions made are, first, that all of the

[1] "Constant Frequency Oscillators," F. B. Llewellyn, *Bell Sys. Tech. Jour.*, January 1932.

[2] "Regeneration Theory," H. Nyquist, *Bell Sys. Tech. Jour.*, January 1932; "Stabilized Feedback Amplifier," H. S. Black, *Bell Sys. Tech. Jour.*, January 1934.

circuit elements except the tube resistances are linear and of lumped character and, second, that modulation effects arising from the nonlinearity of the tube resistances may be neglected. The validity of the second assumption is discussed in the appendix to the article by Llewellyn noted above. Its use permits the treatment of the system as though the resistances were actually linear but variable in magnitude in response to variations of the oscillation amplitude.

Theory

The essential features of the single-tube feedback oscillator are shown in Fig. 1. The feedback network B is unrestricted in its configuration

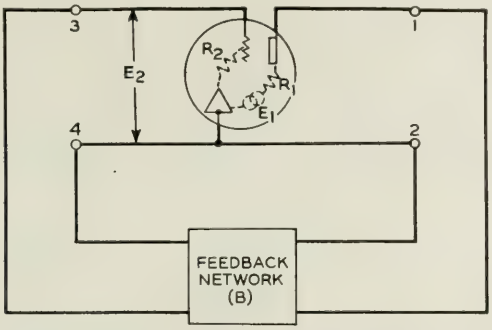


Fig. 1—Elements of a single tube feedback oscillator.

and complexity and may include the vacuum tube electrode capacitances in addition to the external elements. The impedance system of the tube is reduced to the plate and grid resistances R_1 and R_2 with unilateral coupling between them, the latter being indicated by the inclusion of a generator in series with the plate resistance. The voltage E_1 generated in the plate circuit is proportional to the voltage E_2 between the grid and the cathode and, when the system is oscillating, the latter voltage is produced entirely by E_1 as the result of the coupling through the feedback network.

The condition for the existence of self-sustained oscillations is expressed very concisely by the familiar equation

$$\mu\beta = 1, \tag{1}$$

wherein μ and β denote the voltage transfer ratios in the vacuum tube and in the feedback path respectively. The factor μ is the negative of the amplification constant of the tube, the negative sign taking account of the phase reversal inherent in a simple triode. The transfer ratio β represents the ratio of E_2 to E_1 for transmission through the feedback network.

Since the transfer ratios μ and β are both complex quantities, equation (1) expresses the two-fold requirement that the magnitude or modulus of $\mu\beta$ shall be unity and that its phase angle shall be zero. Taking the factors separately, it follows that the modulus of β must be the reciprocal of the modulus of μ and that the phase angles of the two must be equal and of opposite sign. For the single-tube oscillator, the phase angle of β must be 180 degrees since the phase angle of μ has that value.

While the relationships stated above are of simple character, they do not by themselves suffice for the calculation of the oscillation frequency from the constants of the tube and the external circuit. The reason for this is that the values of the tube resistances R_1 and R_2 enter into the determination of the frequency in the general case, and, since these are dependent upon the oscillation amplitude, they cannot be known until the final steady amplitude of the oscillations is known. What happens in an actual oscillator circuit is that, as the oscillation amplitude grows, after initiation, there is a mutual adjustment of frequency and of the resistance values until a condition is reached under which both requirements are met simultaneously. In the case of a stabilized oscillator, since the frequency is independent of the tube resistances, the conditions are simplified and the oscillation frequency can be determined directly by means of the relationships stated above. The non-linearity of the resistances affects only the amplitude of the oscillations.

The evaluation of $\mu\beta$ in terms of the impedance parameters of the circuit permits the determination of the specific circuit conditions in any case for the generation of steady oscillations. The determination is simplified by the consideration that the factor μ has a constant phase angle of 180 degrees so that the variation of the phase of $\mu\beta$ is wholly that of the factor β.

General Formulae for $\mu\beta$

The feedback path, or β circuit, is shown separately in Fig. 2, the

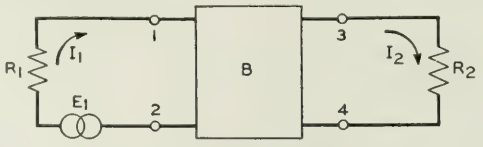


Fig. 2—Simplified schematic of an oscillator feedback circuit.

notations being the same as in Fig. 1. The network B may be of any degree of complexity, but may be assumed to be made up of lumped

impedances. In writing down the equations for the mesh currents, let it be assumed that the circuit contains n meshes including the terminal meshes and that the meshes are so chosen that the resistances R_1 and R_2 do not appear as mutual impedances. Designating the meshes in which R_1 and R_2 appear as the first and second respectively, the mesh current equations take the form

$$\begin{array}{ccc|c} I_1, & I_2, & I_3 \cdots I_n & \\ \hline R_1 + Z_{11}, & Z_{12}, & Z_{13} \cdots Z_{1n} & E \\ Z_{21}, & R_2 + Z_{22}, & Z_{23} \cdots Z_{2n} & 0 \\ Z_{31}, & Z_{32}, & Z_{33} \cdots Z_{3n} & 0 \\ — & — & — \quad — \quad — & — \\ Z_{n1}, & Z_{n2}, & Z_{n3} \cdots Z_{nn} & 0 \end{array} \qquad (2)$$

The subscripts of the Z's denote self and mutual impedances in accordance with the usual conventions, the latter being subject to the reciprocal relationships characteristic of linear systems.

The solution of the above equations for the current I_2 is

$$I_2 = \frac{-E\Delta_{21}}{\Delta + R_1R_2\Delta_{11,\,22} + R_1\Delta_{11} + R_2\Delta_{22}}, \qquad (3)$$

where Δ is the determinant of the coefficients of equations (2) for zero values of R_1 and R_2, and the other determinants are the minors of Δ obtained by crossing out the columns and rows indicated by the numerical subscripts. Thus Δ_{21} is obtained by crossing out the second column and the first row of Δ and $\Delta_{11,\,22}$ is obtained by crossing out the first two columns and the first two rows.

Since, by definition,

$$\beta = \frac{I_2R_2}{E},$$

equation (3) gives

$$\beta = \frac{-R_2\Delta_{21}}{\Delta + R_1R_2\Delta_{11,\,22} + R_1\Delta_{11} + R_2\Delta_{22}}. \qquad (4)$$

The factor μ is the negative of the amplification constant of the tube and if the latter be denoted by α, the value of $\mu\beta$ becomes

$$\mu\beta = \frac{\alpha R_2\Delta_{21}}{\Delta + R_1R_2\Delta_{11,\,22} + R_1\Delta_{11} + R_2\Delta_{22}} \qquad (5)$$

The determinants appearing in equations (4) and (5) can be expanded by the ordinary processes to give expressions in terms of the mesh impedances in any particular case. However, as they stand, they

are significant parameters of the system and, for the present, need no further expansion.

Another general formula for $\mu\beta$ is obtained by making use of the image parameters of the coupling network. If the image impedances at terminals 1, 2, and terminals 3, 4, are denoted by K_1 and K_2, respectively, and the image transfer constant by θ, then

$$\mu\beta = \frac{-\alpha R_2\sqrt{K_1K_2}}{(K_1K_2 + R_1R_2)\sinh\theta + (R_1K_2 + R_2K_1)\cosh\theta}. \tag{6}$$

The two sets of parameters are related by the equations

$$\begin{aligned} K_1^2 &= \frac{\Delta_{22}}{\Delta_{11}} \cdot \frac{\Delta}{\Delta_{11,\,22}}, \\ K_2^2 &= \frac{\Delta_{11}}{\Delta_{22}} \cdot \frac{\Delta}{\Delta_{11,\,22}}, \\ \tanh^2\theta &= \frac{\Delta\Delta_{11,\,22}}{\Delta_{11}\Delta_{22}}. \end{aligned} \tag{7}$$

Equation (6) is useful in many cases because of the fact that the frequency characteristics of the image parameters are well known for a large number of circuit configurations, particularly those of wave filters.

In dealing with many practical oscillator circuits, the simplifying assumption may be made that the coupling network contains only pure reactances. The determinants in equation (5) then become either real quantities or pure imaginaries, thus making it easy to separate the real and the imaginary parts of $\mu\beta$. If the number of meshes in the β circuit, or the number of rows or columns in the determinant Δ, is even, then Δ will be real and if the number is odd Δ will be imaginary. The determinant $\Delta_{11,\,22}$ will be of the same character as Δ, but will take the opposite sign, and determinants Δ_{11}, Δ_{22}, and Δ_{21} will be imaginary when Δ is real and real when Δ is imaginary. Accordingly, equation (5) may be transformed to

$$\mu\beta = \frac{\alpha R_2 D_{21}}{(R_1D_{11} + R_2D_{22}) + j(D - R_1R_2D_{11,\,22})}, \tag{8}$$

in which the D's are determinants of the mesh reactances corresponding respectively to the Δ's of equation (5) having the same subscripts. The phase angle of $\mu\beta$, denoted by φ, is given by

$$\tan\varphi = \frac{D - R_1R_2D_{11,\,22}}{R_1D_{11} + R_2D_{22}} \tag{9}$$

and the value of $\mu\beta$, when tan φ is zero, by

$$\mu\beta_0 = \frac{\alpha R_2 D_{21}}{R_1 D_{11} + R_2 D_{22}}. \tag{10}$$

The angle φ may be either zero or 180 degrees when tan φ is zero, but which it is may be determined by the sign of $\mu\beta_0$. If this is positive, the phase angle is zero, the phase angle of β being then 180 degrees.

For the simplified case of the pure reactance coupling network the expression for $\mu\beta$ in terms of the image parameters takes different forms depending upon whether the frequency lies in a transmission band or in an attenuating band. At frequencies within a transmission band the image impedances are resistive and the transfer constant θ is a pure imaginary quantity indicating a phase shift without attenuation. Denoting this phase shift by ψ, equation (6) becomes

$$\mu\beta = \frac{-\alpha R_2 \sqrt{K_1 K_2}}{(R_1 K_2 + R_2 K_1)\cos\psi + j(K_1 K_2 + R_1 R_2)\sin\psi}, \tag{11}$$

from which

$$\tan\varphi = \frac{(K_1 K_2 + R_1 R_2)}{(R_1 K_2 + R_2 K_1)} \tan\psi. \tag{12}$$

At the cut-off frequencies, equations (11) and (12) become indeterminate. At frequencies in the attenuation bands, the transfer constant θ takes the form

$$\theta = A + j\frac{n\pi}{2}, \tag{13}$$

where A denotes the attenuation and n is an integer the value of which may be different for the different transmission bands of a complex network. The image impedances are pure imaginaries, but their product is real and may be either positive or negative. Simplified forms of equation (6) may be written down for any particular case.

Frequency Stabilization

From equation (9), which gives the value of the phase angle of $\mu\beta$, it is at once evident that an essential condition for zero phase angle is that

$$D - R_1 R_2 D_{11,\,22} = 0. \tag{14}$$

Since the quantities D and $D_{11,\,22}$ are functions of frequency, equation (14) determines the frequency or frequencies at which the phase shift is zero and hence determines the oscillation frequency. The equation

may be satisfied in two distinctly different ways. In accordance with the first, both D and $D_{11,\ 22}$ may be finite and of like sign, in which case the frequency depends upon the resistance product R_1R_2. Since these resistances vary with the oscillation amplitude or with the vacuum tube excitation voltages, a solution of this type is indicative of instability of the frequency.

The second way in which equation (14) may be satisfied depends on the fact that D and $D_{11,\ 22}$ may each have zero values at one or more frequencies according to the degree of complexity of the coupling network. If, then, the network can be so designed that D and $D_{11,\ 22}$ each have a zero at the same frequency and are of opposite sign elsewhere, the condition expressed by the equation will be satisfied at that frequency for any values of the tube resistances and will be satisfied at that frequency only. Whether or not oscillations can be sustained at the frequency so determined may be ascertained readily with the help of equation (10). Oscillations occurring under the above condition are theoretically stable. As demonstrated experimentally by Llewellyn, a very high degree of constancy of the frequency is obtained in actual circuits.

The method of stabilization described above consists in establishing a limited frequency interval within which the oscillation frequency must necessarily lie and then reducing the width of the interval to substantially zero. The establishment of the finite interval is a matter of the choice of an appropriate circuit configuration and the determination of its limits is effected by suitable proportioning of the elements.

Considered in the light of the image parameters of the feedback network, the method of stabilization consists in making the image phase angle of the network take the value 180 degrees at a frequency within a transmission band. Referring to equations (11) and (12), it will be seen that when ψ, the image phase angle of the network, takes the value 180 degrees, the quantity $\mu\beta$ becomes real and positive, indicating the possibility of self-oscillation. Since the result is independent of the values of the tube resistances, the oscillations are theoretically stable.

In an attenuation band, the transfer constant θ may include a phase angle of 180 degrees which is constant with frequency, but, because of the real component representing attenuation, neither cosh θ nor sinh θ in equation (6) can become zero at any frequency. To obtain an overall phase shift of 180 degrees in the feedback path it is therefore necessary that

$$K_1K_2 + R_1R_2 = 0, \tag{15}$$

which requires the product K_1K_2 to be negative and, hence, that both image impedances be reactances of the same sign. Since the condition expressed by equation (15) is dependent upon the values of the tube resistances, it follows that frequency stabilization cannot be obtained at frequencies in an attenuation band.

The problem of devising stabilized reactance type oscillator circuits therefore resolves itself into that of obtaining band-pass coupling networks which have phase constants of 180 degrees at frequencies within the pass-bands. Evidently there is a multiplicity of known filter structures that meet the requirements and also many other networks of similar character including all-pass reactance networks. Since each half-section of a filter gives a phase shift of 90 degrees in the pass-band, it follows that the coupling network should be equivalent to at least three half-sections. More complex networks may be used, but with networks equivalent to more than six half-sections oscillations may occur at more than one frequency.

Illustrative Examples

The principles discussed in the foregoing section will be illustrated by a consideration of the circuit shown in Fig. 3, in which the coupling

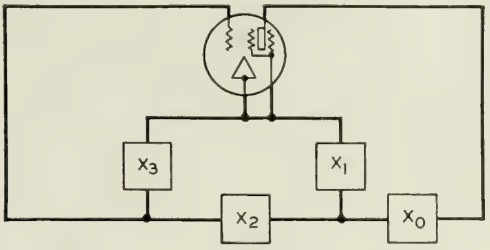


Fig. 3—Oscillator with plate circuit stabilizing impedance.

network is a simple ladder network having four reactive branches. A screen-grid vacuum tube is assumed so that the grid-to-plate capacitance is negligibly small and all feedback is confined to the coupling network. For this circuit, the determinant D has the value

$$D = X_3(X_0X_1 + X_1X_2 + X_2X_0). \tag{16}$$

The several minors are

$$D_{11,\,22} = X_1 + X_2 + X_3, \tag{17}$$

$$D_{11} = X_3(X_1 + X_2), \tag{18}$$

$$D_{22} = X_0(X_1 + X_2 + X_3) + X_1(X_2 + X_3) \tag{19}$$

and

$$D_{21} = -X_1X_3. \tag{20}$$

The condition that D and $D_{11,\,22}$ be zero at the same frequency might be met by making the reactance X_3 a simple series resonant combination and proportioning X_1 and X_2 to be resonant at the same frequency as X_3. The two determinants would then both have zeros at this frequency, but oscillations could not occur since D_{21} would also become zero and the feedback would be destroyed. The necessary condition for stabilization is, therefore, that $(X_1 + X_2 + X_3)$ and $(X_0X_1 + X_1X_2 + X_2X_0)$ become zero at the same frequency. It is readily shown that this can be achieved by making the impedance X_0 such that

$$X_0X_3 = X_1X_2 \tag{21}$$

at the frequency for which $(X_1 + X_2 + X_3)$ is zero. The addition of the plate circuit reactance X_0 provides a circuit configuration which makes stabilization possible. The character of the several branch reactances may be such that equation (21) holds at all frequencies but it need hold only at zero of $D_{11,\,22}$. This minimum restriction permits considerable diversification of the form of the coupling network.

The modified Colpitts oscillator shown in Fig. 4 is a simple case of the general circuit of Fig. 3. For this circuit the reactance determinants have the values

$$D = -\frac{L_0L_2}{\omega C_3}\left(\omega^2 - \frac{1}{L_2C_1} - \frac{1}{L_0C_1}\right), \tag{22}$$

$$D_{11,\,22} = \frac{L_2}{\omega}\left(\omega^2 - \frac{1}{L_2C_1} - \frac{1}{L_2C_3}\right), \tag{23}$$

$$D_{11} = -\frac{L_2}{\omega^2C_3}\left(\omega^2 - \frac{1}{L_2C_1}\right), \tag{24}$$

$$D_{22} = L_0L_2\left(\omega^2 - \frac{1}{L_2C_1} - \frac{1}{L_2C_3}\right) - \frac{L_2}{\omega^2C_1}\left(\omega^2 - \frac{1}{L_2C_3}\right), \tag{25}$$

$$D_{21} = -\frac{1}{\omega^2C_1C_3}. \tag{26}$$

Both D and $D_{11,\,22}$ have frequency variations corresponding to those of simple resonant circuits but, in the case of the former, with the sign reversed. The two quantities have the same sign only in the interval between the two resonance frequencies or zeros and, since these resonance frequencies are independently adjustable, the interval may be made as small as may be desired. The interval is reduced to

zero and the oscillation frequency stabilized when the elements are so proportioned that

$$L_0C_1 = L_2C_3, \tag{27}$$

under which condition the oscillation frequency is determined by the equation

$$\omega_0^2 = \frac{1}{L_2}\left(\frac{1}{C_1} + \frac{1}{C_3}\right). \tag{28}$$

At the oscillation frequency the value of $\mu\beta$ (equation 10) becomes

$$\mu\beta_0 = \frac{\alpha R_2}{R_1\dfrac{C_1}{C_3} + R_2\dfrac{C_3}{C_1}} \tag{29}$$

and is equal to unity when

$$\frac{R_1}{R_2} = \frac{C_3}{C_1}\left(\alpha - \frac{C_3}{C_1}\right). \tag{30}$$

Equation (30) can be used to determine the amplitude of the stabilized oscillations if the variation of the resistance ratio with amplitude is known or can be found experimentally. At the moment of inception of the oscillations the amplitude will be infinitesimally small and the tube resistances will generally be such that the initial value of $\mu\beta$ is considerably greater than unity. As the amplitude grows, the plate resistance R_1 tends to increase and the grid resistance to diminish until at a certain amplitude the resistance ratio takes the value given by equation (25). The oscillations then remain steady at this amplitude.

It may be noted that the amplitude relationship (29) holds so long as the capacitances are maintained in the fixed ratio

$$\frac{C_3}{C_1} = \frac{L_0}{L_2} \tag{31}$$

and is independent of their absolute values. If, therefore, the capacitances are varied simultaneously while their ratio is maintained constant, the oscillation frequency will be varied, but frequency stability will be maintained for all adjustments and the oscillation amplitude will remain constant. A similar result may be obtained by varying the inductances simultaneously.

It is instructive to examine the action of the added plate circuit inductance in Fig. 4 in the light of the image parameters of the coupling

network. When the values of the inductances and capacitances are unrestricted the coupling network will, in general, have three pass-bands, each pass-band being characterized by a purely imaginary value

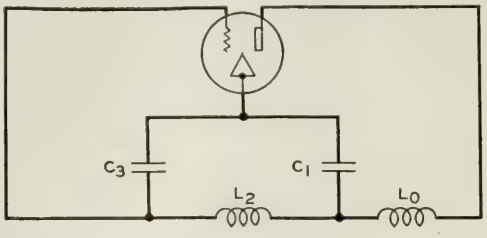


Fig. 4—Colpitts oscillator with plate circuit stabilization.

of the image transfer constant θ corresponding to a phase shift without attenuation. The phase shift is zero at zero frequency, increases by 90 degrees in each pass-band, and remains constant at 90, 180, 270 degrees in the successive attenuation bands. From the expression for tanh θ in equation (7) it follows that the image phase shift of the reactive network will be zero or an even multiple of 90 degrees at the zeros of D and $D_{11,\ 22}$ and will be an odd multiple of 90 degrees at the zeros of D_{11} and D_{22}. The general phase shift characteristic of the feedback network in Fig. 4 is shown in Fig. 5. The critical frequencies f_3 and f_4

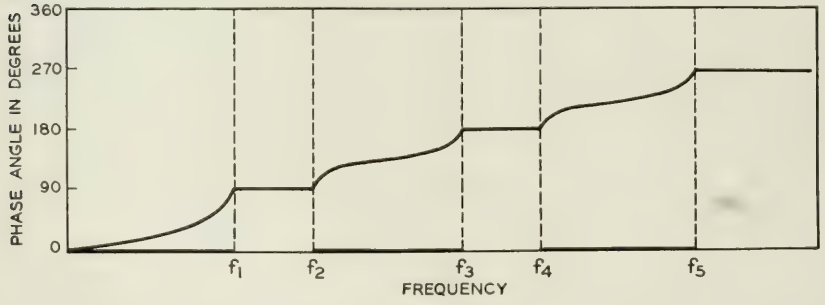


Fig. 5—Phase constant characteristic of the feedback network in an incompletely stabilized Colpitts oscillator.

marking the edges of the second attenuation band correspond to the zeros of D and $D_{11,\ 22}$. Frequencies f_2 and f_5 are the zeros of D_{22} and f_1 is the single zero of D_{11}. The pass-bands are indicated by the heavy lines on the frequency scale. An examination of the image impedances of the network will show that in the second attenuation band they are reactances of like sign and are of opposite sign in the first attenuation band.

The necessary condition for self-oscillation stated in equation (15) is satisfied in the attenuation band between f_3 and f_4. The overall phase shift in the feedback path will follow the image phase shift characteristic approximately and will be equal to 180 degrees at some frequency in this range depending upon the values of the terminal resistances provided by the tube space paths. Oscillations may result but their frequency will be unstable. By reducing the width of the attenuation band stability is increased and becomes theoretically complete when the band is reduced to zero. Proportioning the circuit in accordance with equation (27) to give stability makes the two upper pass-bands of the network confluent. If the whole network be proportioned as a one-and-a-half-section constant-k filter all three pass-bands become confluent.

Since the stabilization requirement

$$X_0X_3 = X_1X_2$$

need hold only at the oscillation frequency, various possible modifications of the circuit of Fig. 4 become readily apparent. For example, the inductance L_2 may be replaced by a series resonant combination which has the same reactance at the oscillation frequency ω_0. This

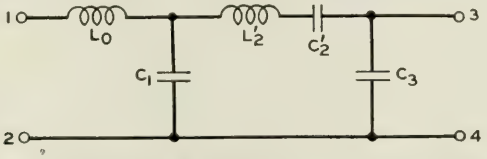


Fig. 6—Modification of the feedback network of the stabilized Colpitts oscillator by the introduction of an extra element.

gives the circuit shown in Fig. 6, in which L_2 is replaced by the combination L_2', C_2' such that

$$L_2' = L_2 + \frac{1}{\omega_0^2 C_2'}. \tag{32}$$

A further possible modification is shown in Fig. 7 in which L_2 is replaced by a three-element combination comprising a series resonant

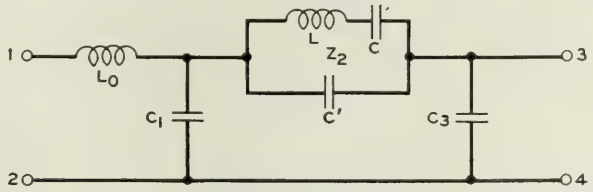


Fig. 7—Further modification of the feedback network of the stabilized Colpitts oscillator.

circuit shunted by the capacitances. At the frequency for which $(X_1 + X_2 + X_3)$ is zero the three-element combination has, as inductive reactance, jX_{20}, which can be computed. The inductance L_0 should then be such that

$$L_0 = \frac{X_{20}}{\omega_0} \cdot \frac{C_3}{C_1}. \tag{33}$$

Evidently the three-element combination in the Z_2 branch is such that it might be replaced by a piezoelectric crystal. To keep the inductance L_0 small, the capacitances C_1 and C_3 should be fairly large so that the resonance of $(X_1 + X_2 + X_3)$ lies close to the crystal resonance.

The foregoing examples are based on the constant-k low-pass filter as a prototype. Evidently high-pass or band-pass filters of the various known kinds might also be used as prototypes and diversified in similar manner. Additional forms may also be found by increasing the number of meshes in the network, but in such cases, the simplest circuits providing frequency stability appear to be the homogeneous single pass-band filter networks. The stable oscillation frequency is the frequency within the pass-band for which the phase constant is equal to 180 degrees. If the network is equivalent to six ladder-type half-sections or more, there will be two or more frequencies for which the phase constant is 180 degrees. Such networks are generally not well-suited for oscillator circuits.

Certain simple configurations which do not admit of complete stabilization in the above manner may be partially stabilized and in actual use may exhibit a very high degree of constancy. The common quartz crystal controlled oscillator with the crystal connected between the grid and cathode of the tube is an example of a partially stabilized circuit. The impedance characteristic of the crystal itself is primarily responsible for the stabilization. Usually the circuit is such as to require the crystal to exhibit an inductive reactance at the oscillation frequency and the impedance characteristic is such that this occurs only in an extremely small frequency interval. The main determinant D for this circuit has a single zero at the resonance of the crystal and complete stabilization would require that the minor $D_{11,22}$ have its zero at this frequency also. However, oscillation under this condition would be impossible since the crystal resonance would short-circuit the feedback path and reduce the magnitude of $\mu\beta$ to zero. Actually the zero of $D_{11,22}$ lies somewhere in the inductive interval of the crystal at a point fairly close to the resonance frequency. The range in which oscillation is possible is therefore only a fraction of the inductive interval of the crystal and a high degree of stability results.

Dissipative Feedback Networks

The foregoing sections deal with non-dissipative feedback networks, but the general ideas set forth are applicable to certain types, at least, of dissipative networks. For such networks the transfer constant θ is a complex quantity and may be represented by

$$\theta = A + j\psi, \tag{34}$$

where A denotes the attenuation and ψ the phase constant. When the phase constant is equal to 180 degrees the hyperbolic functions of the transfer constant take the values

$$\begin{aligned} \sinh \theta &= -\sinh A, \\ \cosh \theta &= -\cosh A, \\ \text{and} \\ \tanh \theta &= \tanh A. \end{aligned} \tag{35}$$

The image impedances will generally be complex, but, if they can be made to become purely resistive at the frequency for which the phase constant is 180 degrees, the value of $\mu\beta$ at that frequency then becomes

$$\mu\beta = \frac{\alpha R_2 \sqrt{\rho_1 \rho_2}}{(\rho_1\rho_2 + R_1R_2)\sinh A + (R_1\rho_2 + R_2\rho_1)\cosh A}, \tag{36}$$

in which ρ_1 and ρ_2 denote the resistive values of K_1 and K_2. Since this is necessarily a positive real quantity the phase angle of $\mu\beta$ is zero and remains zero independently of variations of R_1 and R_2. Oscillations occurring under this condition are therefore theoretically stable.

Two Tube Oscillators

The stabilization of the single-tube oscillator depends on the circumstances that the tube itself produces a constant phase shift of 180 degrees and that feedback networks can be devised to produce a phase shift of this value which is independent of the terminal resistances. Phase shifts of 90 degrees which are independent of the termination can also be provided by means of reactive networks and this property may likewise be made use of in the design of stabilized oscillators. For this purpose it is necessary to have an amplifier which will give a uniform phase shift of 90 degrees over a fairly wide range of frequencies in the neighborhood of the oscillation frequency. A suitable amplifier may consist of two vacuum tubes coupled in tandem by a simple shunt inductance or a simple shunt capacitance. The second tube should be

suitably biased to avoid drawing grid current during operation and the coupling reactance should be small in comparison with the plate resistance of the first tube. Preferably the first tube should be of the screen-grid type having a mutual conductance which is independent of the connected output impedance. The two tubes by themselves have a total phase shift of 360 degrees or zero, but the shunt coupling reactance in combination with the internal resistance of the first tube provides a further phase shift of 90 degrees which represents the total effective phase shift of the amplifier.

With a phase shift of 90 degrees in the amplifier, the oscillation frequency will be that for which the feedback path has a phase shift of 90 degrees in the reverse direction. The conditions for frequency stabilization follow readily from the principles already developed.

For a purely reactive feedback network, the expression for $\mu\beta$ may be obtained from equation (8) by substituting $\pm j\alpha$ for the amplification factor. This gives

$$\mu\beta = \frac{\pm j\alpha R_2 D_{21}}{(R_1 D_{11} + R_2 D_{22}) + j(D - R_1 R_2 D_{11,\ 22})}, \tag{37}$$

from which is obtained

$$\tan \varphi = \pm \frac{R_1 D_{11} + R_2 D_{22}}{D - R_1 R_2 D_{11,\ 22}}, \tag{38}$$

and

$$\mu\beta_0 = \pm \frac{\alpha R_2 D_{21}}{D - R_1 R_2 D_{11,\ 22}}. \tag{39}$$

From equation (39) giving the phase angle of $\mu\beta$, it is evident that the phase angle can be zero independently of the magnitudes of the tube resistances only if D_{11} and D_{22} have zero values at a common frequency. This then is the criterion for stability of the oscillation frequency.

In terms of the image parameters of the coupling networks, the value of $\mu\beta$ becomes

$$\mu\beta = \frac{\pm j\alpha R_2 \sqrt{K_1 K_2}}{(K_1 K_2 + R_1 R_2) \sinh \theta + (R_1 K_2 + R_2 K_1) \cosh \theta}. \tag{40}$$

If it be assumed that the network is dissipative, the transfer constant has an attenuation component and may be represented by

$$\theta = A + j\psi$$

as in equation (34). When the phase constant ψ has the value $\pm$ 90 degrees, equation (40) becomes

$$\mu\beta = \frac{\pm \alpha R_2 \sqrt{K_1 K_2}}{(K_1K_2 + R_1R_2) \sinh A + (R_1K_2 + R_2K_1) \cosh A}. \quad (41)$$

When the feedback network is purely reactive, stabilization of the frequency requires that the phase shift component of the image transfer constant have a value $\pm$ 90 degrees within a transmission band. Under this condition the attenuation component of the transfer constant is zero and equation (41) is simplified by the reduction of sinh A to zero and cosh A to unity.

A simple example of an oscillator stabilized in the above manner is shown in Fig. 8. The two vacuum tubes are coupled by a simple shunt

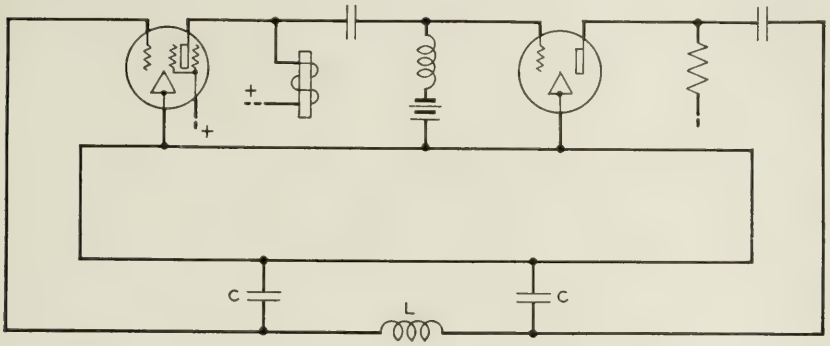


Fig. 8—Stabilized two-stage oscillator.

inductance of relatively low magnitude and low dissipation. With the high resistance of the screen-grid tube in the first stage this shunt reactance coupling provides a substantially constant phase shift of 90 degrees. A low-impedance transformer might also be used for coupling the stages or, if desired, a four-terminal dissipative network designed to provide the required phase shift in a moderately wide frequency range.

The feedback network is required to produce a phase shift of only 90 degrees and may therefore have a relatively simple configuration. The direction of the phase shift should, of course, be opposite to that of the amplifier. In the example illustrated the feedback network corresponds to that of a simple Colpitts oscillator. The condition for stabilization is that the two shunt capacitances be equal and the oscillation frequency is that of the resonance of the inductance with one of the two equal condensers.

It will be evident from the foregoing that a very large number of circuit configurations of the feedback network will provide theoretical stabilization of the oscillation frequency either with single stage or with multiple amplifiers. Naturally, not all of these will be of the same practical interest, since an undue complexity of the network may give rise to difficulties in its adjustment and may increase the problem of temperature compensation.

The Discovery of Electron Waves *

By C. J. DAVISSON

THAT streams of electrons possess the properties of beams of waves was discovered early in 1927 in a large industrial laboratory in the midst of a great city, and in a small university laboratory overlooking a cold and desolate sea. The coincidence seems the more striking when one remembers that facilities for making this discovery had been in constant use in laboratories throughout the world for more than a quarter of a century. And yet the coincidence was not, in fact, in any way remarkable. Discoveries in physics are made when the time for making them is ripe, and not before; the stage is set, the time is ripe and the event occurs—more often than not at widely separated places at almost the same moment.

The setting of the stage for the discovery of electron diffraction was begun, one may say, by Galileo. But I do not propose to emulate the gentleman who began a history of his native village with the happenings in the Garden of Eden. I will take, as a convenient starting point, the events which led to the final acceptance by physicists of the idea that light for certain purposes must be regarded as corpuscular. This idea after receiving its quietus at the hands of Thomas Young in 1800 return to plague a complacent world of physics in the year 1899. In this year Max Planck put forward his conception that the energy of light is in some way quantized. A conception which, if accepted, supplied, as he showed, a means of explaining completely the distribution of energy in the spectrum of black body radiation. The quantization was such that transfers of energy between radiation and matter occurred abruptly in amounts proportional to the radiation frequency. The factor of proportionality between these quantities is the ever-recurring Planck constant, h. Thus was reborn the idea that light is in some sense corpuscular.

How readily this circumstantial evidence for a corpuscular aspect of light would have been accepted as conclusive must remain a matter of conjecture, for already the first bits of direct evidence pointing to the same conclusion were being taken down from the scales and meters of

* Nobel lecture delivered at Stockholm, Sweden, December 13, 1937 in connection with the award of the 1937 Nobel Prize in Physics by the Swedish Academy of Science to Dr. Davisson and Professor George P. Thompson of London. This lecture was originally published in the year book Les Prix Nobel, 1937.

the laboratory; the truth about light was being wrung from Nature—at times, and in this case, a most reluctant witness.

In an extended examination carried on chiefly by Richardson and K. T. Compton, Hughes, and Millikan, it was brought out that light imparts energy to individual electrons in amounts proportional to its frequency and finally that the factor of proportionality between energy and frequency is just that previously deduced by Planck from the black body spectrum. The idea of pressing the witness on the latter point had come from Einstein who out-plancked Planck in not only accepting quantization, but in conceiving of light quanta as actual small packets or particles of energy transferable to single electrons in toto.

The case for a corpuscular aspect of light, now exceedingly strong, became overwhelmingly so when in 1922 A. H. Compton showed that in certain circumstances light quanta—photons as they were now called—have elastic collisions with electrons in accordance with the simple laws of particle dynamics. What appeared, and what still appears to many of us as a contradiction in terms had been proved true beyond the least possible doubt—light was at once a flight of particles and a propagation of waves; for light persisted, unreasonably, to exhibit the phenomenon of interference.

Troubles, it is said, never come singly, and the trials of the physicist in the early years of this century give grounds for credence in the pessimistic saying. Not only had light, the perfect child of physics, been changed into a gnome with two heads—there was trouble also with electrons. In the open they behaved with admirable decorum, observing without protest all the rules of etiquette set down in Lorentz's manual, but in the privacy of the atom they indulged in strange and unnatural practices; they oscillated in ways which no well-behaved mechanical system would deem proper. What was to be said of particles which were ignorant apparently of even the rudiments of dynamics? Who could apologize for such perversity—rationalize the data of spectroscopy? A genius was called for, and a genius appeared. In 1913 Niels Bohr gave us his strange conception of "stationary" orbits in which electrons rotated endlessly without radiating, of electrons disappearing from one orbit and reappearing, after brief but unexplained absences, in another. It was a weird picture—a picture to delight a Surrealist—but one which fascinated the beholder, for in it were portrayed with remarkable fidelity the most salient of the orderly features which spectroscopic data were then known to possess; there was the Balmer series! and there the Rydberg constant!—correct to the last significant digit! It was a masterpiece. It is important to

note that in achieving this *tour de force* Bohr made judicious use of the constant which Planck had extracted from the black body spectrum, the constant h.

It looked at this time—in the year 1913—as if the authentic key to the spectra had at last been found, as if only time and patience would be needed to resolve their riddles completely. But this hope was never fulfilled. The first brilliant triumphs of the theory were followed by yet others, but soon the going became distressingly difficult, and finally, despite the untiring efforts of countless helpers, the attack came virtually to a standstill. The feeling grew that deeply as Bohr had dived he had not, so to speak, touched bottom. What was wanted, it was felt, was a new approach, a new theory of the atom which would embrace necessarily all the virtues of the Bohr theory and go beyond it —a theory which would contain some vaguely sensed unifying principle which, it was felt, the Bohr theory lacked.

Such an underlying principle had been sought for almost from the first. By 1924 one or two ideas of promise had been put forward and were being assiduously developed. Then appeared the brilliant idea which was destined to grow into that marvelous synthesis, the present day quantum mechanics. Louis de Broglie put forward in his doctor's thesis the idea that even as light, so matter has a duality of aspects; that matter like light possesses both the properties of waves and the properties of particles. The various "restrictions" of the Bohr theory were viewed as conditions for the formation of standing electron wave patterns within the atom.

Reasoning by analogy from the situation in optics and aided by the clue that Planck's constant is a necessary ingredient of the Bohr theory, de Broglie assumed that this constant would connect also the particle and wave aspects of electrons, if the latter really existed. De Broglie assumed that, as with light, the correlation of the particle and wave properties of matter would be expressed by the relations:

(Energy of particle) $E = h\nu$ (frequency, waves/unit time).
(Momentum of particle) $p = h\sigma$ (wave number, waves/unit distance).

The latter may be written in the more familiar form $\lambda = h/p$ where λ represents wave-length.

Perhaps no idea in physics has received so rapid or so intensive development as this one. De Broglie himself was in the van of this development but the chief contributions were made by the older and more experienced Schroedinger.

In these early days—eleven or twelve years ago—attention was focussed on electron waves in atoms. The wave mechanics had

sprung from the atom, so to speak, and it was natural that the first applications should be to the atom. No thought was given at this time, it appears, to electrons in free flight. It was implicit in the theory that beams of electrons like beams of light would exhibit the properties of waves, that scattered by an appropriate grating they would exhibit diffraction, yet none of the chief theorists mentioned this interesting corollary. The first to draw attention to it was Elsasser, who pointed out in 1925 that a demonstration of diffraction would establish the physical existence of electron waves. The setting of the stage for the discovery of electron diffraction was now complete.

It would be pleasant to tell you that no sooner had Elsasser's suggestion appeared than the experiments were begun in New York which resulted in a demonstration of electron diffraction—pleasanter still to say that the work was begun the day after copies of de Broglie's thesis reached America. The true story contains less of perspicacity and more of chance. The work actually began in 1919 with the accidental discovery that the energy spectrum of secondary electron emission has, as its upper limit, the energy of the primary electrons, even for primaries accelerated through hundreds of volts; that there is, in fact, an elastic scattering of electrons by metals.

Out of this grew an investigation of the distribution-in-angle of these elastically scattered electrons. And then chance again intervened; it was discovered, purely by accident, that the intensity of elastic scattering varies with the orientations of the scattering crystals. Out of this grew, quite naturally, an investigation of elastic scattering by a single crystal of predetermined orientation. The initiation of this phase of the work occurred in 1925, the year following the publication of de Broglie's thesis, the year preceding the first great developments in the wave mechanics. Thus the New York experiment was not at its inception, a test of the wave theory. Only in the summer of 1926, after I had discussed the investigation in England with Richardson, Born, Franck and others, did it take on this character.

The search for diffraction beams was begun in the autumn of 1926, but not until early in the following year were any found—first one and then twenty others in rapid succession. Nineteen of these could be used to check the relationship between wave-length and momentum and in every case the correctness of the de Broglie formula, $\lambda = h/p$ was verified to within the limit of accuracy of the measurements.

I will recall briefly the scheme of the experiment. A beam of electrons of predetermined speed was directed against a (111) face of a crystal of nickel as indicated schematically in Fig. 1. A collector designed to accept only elastically scattered electrons and their near

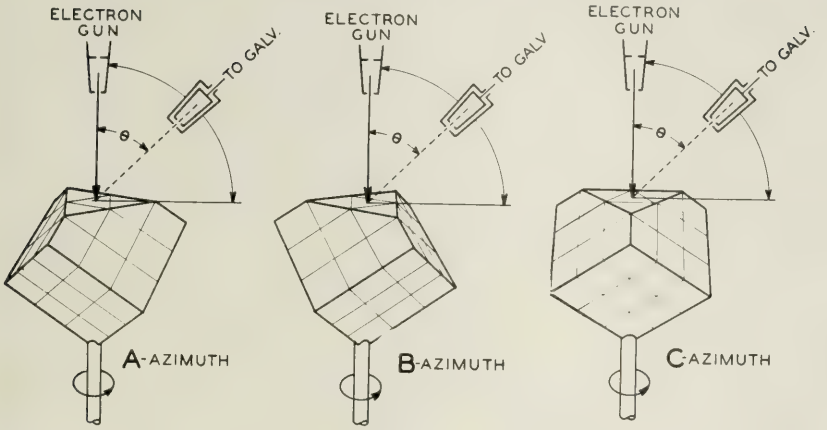


Fig. 1—Schematic diagram showing disposition of primary beam, nickel crystal and collector. Crystal shown revolved to bring one principal azimuth after another into plane of observation.

neighbors, could be moved on an arc about the crystal. The crystal itself could be revolved about the axis of the incident beam. It was possible thus to measure the intensity of elastic scattering in any direction in front of the crystal face with the exception of those directions lying within 10 or 15 degrees of the primary beam.

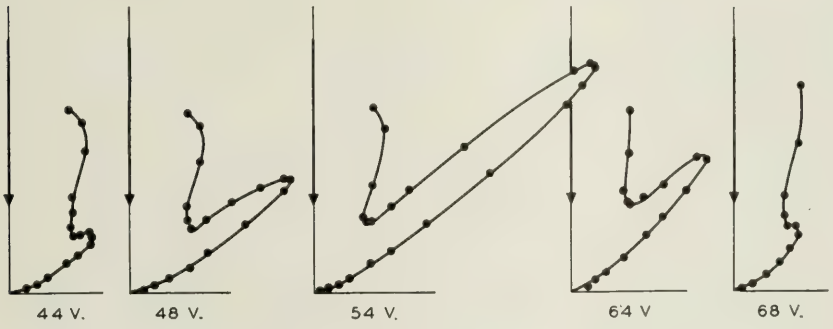


Fig. 2—Polar diagram showing intensity of elastic scattering in A-azimuth (see Fig. 1) as function of latitude angle, for series of primary beam voltages.

The curves reproduced in Fig. 2 show the distribution-in-angle of intensity for a particular azimuth of the crystal. The curves are for a series of electron speeds, therefore, for a series of electron wave-lengths. For a particular wave-length a diffraction beam shines out. Setting the collector on this beam at its brightest and revolving the crystal, the intensity was found to vary in azimuth as illustrated in Fig. 3.

The high peak on the left represents the cross-section-in-azimuth of the beam shown in Fig. 2. Two similar peaks mark the positions of companion beams which with the first form a set of three, as required by the threefold symmetry of the crystal about its (111) directions—the direction of the incident beam. The lesser intermediate peaks are due to a different set of beams which is not here fully developed.

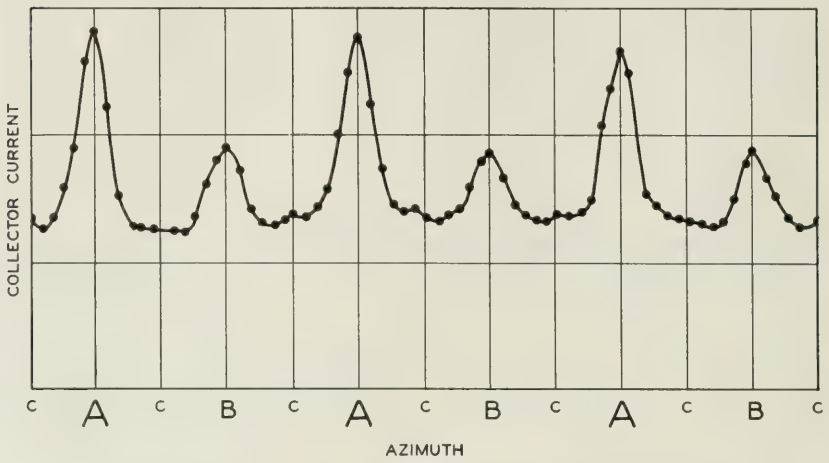


Fig. 3—Curve showing intensity of elastic scattering of 54-volt primary beam as function of azimuth for latitude of peak in 54-volt curve, Fig. 2.

The de Broglie relation was tested by computing wave-lengths from the angles of the diffraction beams and the known constant of the crystal, and comparing these with corresponding wave-lengths computed from the formula $\lambda = h/p$, where p, the momentum of the electrons, was obtained from the potential used to accelerate the beam and the known value of e/m for electrons. If wave-lengths computed from the formula agreed with those obtained from the diffraction data, the de Broglie relation would be verified. How nearly the theoretical values agreed with the experimental is illustrated in Fig. 4. For perfect agreement all points would fall on the line drawn through the origin.

You will realize without my telling you that this series of experiments extending in time over a period of eight or nine years and requiring the construction and manipulation of intricate apparatus was not made by me alone. From first to last a considerable number of my colleagues contributed to the investigation. Chief among these were my two exceptionally able collaborators, Dr. C. H. Kunsman and Dr. L. H. Germer. Dr. Kunsman worked with me throughout the early stages of the investigation, and Dr. Germer, to whose skill and perseverance

a great part of the success of the definitive experiments is due, succeeded Dr. Kunsman in 1924.

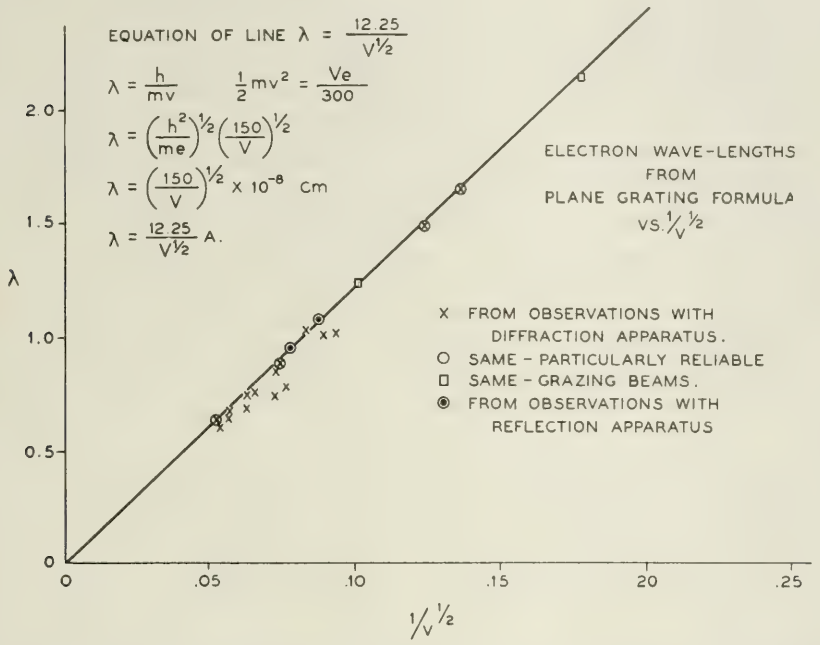


Fig. 4—Test of the de Broglie formula $\lambda = h/p = h/mv$. Wave-length computed from diffraction data plotted against $1/V^{1/2}$, (V, primary beam voltage). For precise verification of the formula all points should fall on the line $\lambda = 12.25/V^{1/2}$ plotted in the diagram.

I would like also at this time to express my admiration of the late Dr. H. D. Arnold, then Director of Research in the Bell Telephone Laboratories, and of Dr. W. Wilson, my immediate superior, who were sufficiently far-sighted to see in these researches a contribution to the science of communication. Their vision was, in fact, accurate for today in ours, as in other industrial laboratories, electron diffraction is applied with great power and efficacy for discerning the structures of materials.

But neither of this nor of the many beautiful and important researches which have been made in electron diffraction in laboratories in all parts of the world since 1927 will I speak today. I will take time only to express my admiration of the beautiful experiments—differing from ours in every respect—by which Thomson in far-away Aberdeen also demonstrated electron diffraction and verified de Broglie's formula at the same time as we in New York. And to mention, as

closely related to the subject of this discourse, the difficult and beautifully executed experiments by which Stern and Esterman in 1929 showed that atomic hydrogen also is diffracted in accordance with the de Broglie-Schroedinger theory.

Important and timely as was the discovery of electron diffraction in inspiring confidence in the physical reality of material waves, our confidence in this regard would hardly be less today, one imagines, were diffraction yet to be discovered, so great has been the success of the mechanics built upon the conception of such waves in clarifying the phenomena of atomic and subatomic physics.

Abstracts of Technical Articles from Bell System Sources

Stability of Two-Meter Waves.[1] CHARLES R. BURROWS, A. DECINO and LOYD E. HUNT. The continuous records of the field strength received over a 60-kilometer path on a frequency of 150 megacycles for the year 1936 are analyzed. Preliminary comparison with other paths of the same length indicate that the magnitude of the recorded variations of the signals may be typical of paths of this length.

A reduction in the path length by a factor of two reduced the fading range in decibels by a factor of five.

The results are found to be in agreement with an earlier formula. Fading reduced the field 7 decibels below the average value 1 per cent of the time.

Loudness, Masking and Their Relation to the Hearing Process and the Problem of Noise Measurement.[2] HARVEY FLETCHER. It is shown in this paper how to define loudness and loudness level in a quantitative way. Definite procedures are given for determining experimentally the loudness level of any sound heard by any person. For a typical observer a true loudness scale is developed. The relation of the scale to the loudness level scale is determined experimentally. The scale has been found to be very useful for calculating loudness from the noise spectrogram, the noise audiogram, or the overtone structure of the sound.

The relation between the masking and the loudness produced by a sound has been quantitatively determined and a formula deduced from this relation which has proved useful for calculating the loudness. This formula may be applied with equal success to a normal ear and also to a deafened ear. Evidence has been given that the masking expressed in decibels produced upon any pure tone is equal directly to the agitation of 1.1 per cent of the total nerve endings expressed in decibels above the threshold value for such a patch and at the position where such a tone would be sensed. These loudness relations throw light upon some of the important processes involved in hearing. In particular the data from the masking effects of thermal noise were used to calculate the relation between the position of

[1] *Proc. I. R. E.*, May 1938.
[2] *Jour. Acous. Soc. Amer.*, April 1938.

maximum stimulation on the basilar membrane and the frequency of the tone producing the stimulation.

Pick-up for Sound Motion Pictures (Including Stereophonic).[3] J. P. MAXFIELD, A. W. COLLEDGE and R. T. FRIEBUS. Although the basic principles underlying sound pick-up for motion pictures have been understood for some time, the ability to carry them out completely in the presence of the requirements of artistry, photography, lighting, etc., has constituted a difficult problem. The paper discusses some of these problems, particularly with respect to the acoustics of production sets and scoring stages. The problems of stereophonic reproduction are also discussed in some detail.

Practical Application of Telephone Repeaters and Carrier Telephone Systems.[4] J. A. PARROTT. The paper discusses engineering problems in the application of telephone repeaters and carrier systems with which railroad communication engineers recently have been particularly concerned. The first part of the paper deals with crosstalk, noise, balance and overloading considerations in the design of repeatered circuits, particularly from the standpoint of selecting the locations of repeaters to obtain the most satisfactory results on existing lines. The importance of securing test data on the wire facilities to aid in this design work as well as to serve as a guide in improving circuit conditions is emphasized.

The second part of the paper briefly discusses the application of the H1 carrier telephone system and provides transmission data for the preliminary design of the layout of such systems. The Type D and K10 carrier transpositions are described and features of particular interest in their possible use on railroad facilities are discussed.

Sorption of Water by Rubber.[5] R. L. TAYLOR and A. R. KEMP. The effect of several variables on the rate of sorption of water by rubber is discussed. Expressions based on short-time immersion tests are derived which permit calculation of the water content after an extended period of immersion under fixed conditions of temperature and vapor pressure. A sorption coefficient by which one material may be compared with another is suggested, and its application to practical problems is considered.

[3] *Jour. S. M. P. E.*, June 1938.
[4] *Proc. Assoc. of Amer. Railroads, Telegraph and Telephone Sec.*, October 1937.
[5] *Indus. & Engg. Chemistry*, April 1938.

Chemical Studies of Wood Preservation—The Wood-Block Method of Toxicity Assay.[6] ROBERT E. WATERMAN, JOHN LEUTRITZ and CALEB M. HILL. Actual decay resistance of treated wood is used as the basis for a simple laboratory technic in the assay of materials advocated for the protection of wood. In its present stage of development the test is a valuable tool in wood preservation studies.

[6] *Indus. & Engg. Chem., Anal. Ed.*, June 15, 1938.

Contributors to this Issue

JULIAN BLANCHARD, A.B., Trinity College (now Duke University), 1905; A.M., Columbia University, 1909; Ph.D., 1917. Professor of Engineering, Trinity College, 1909–1912; Research Assistant in Physics, Columbia University, 1912–1915. Physicist, Research Laboratory, Eastman Kodak Company, 1915–1917; Engineering Department, Western Electric Company, 1917–1925; Bell Telephone Laboratories, 1925–. Dr. Blanchard's work has been concerned primarily with special studies in connection with the development of vacuum tubes and radio.

B. L. CLARKE, B.S., George Washington University, 1921; M.A., Columbia University, 1923; Ph.D., Columbia University, 1924. Bell Telephone Laboratories, 1927–. Dr. Clarke has been in charge of the work in analytical chemistry since 1930.

C. J. DAVISSON, B.Sc., University of Chicago, 1908; Ph.D., Princeton University, 1911; Instructor in Physics, Carnegie Institute of Technology, 1911–17. Engineering Department of the Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. As Research Physicist, Dr. Davisson is engaged in work relating largely to thermionics and electronic physics.

In 1928 the National Academy of Sciences awarded the Comstock Prize to Dr. Davisson "for the most important discovery of or investigation in electricity or magnetism or radiant energy" made in this country during the preceding five years, for his work in this field. In 1931 he and Dr. L. H. Germer received the Elliott Cresson Medals from the Franklin Institute, Philadelphia, and in 1935 he received the Hughes Medal of the Royal Society of London.

W. G. GUSTAFSON, B.S. in Electrical Engineering, Union College, 1927; Columbia University, 1929–36. Bell Telephone Laboratories, 1927–. Mr. Gustafson is engaged in work relating to the development of transformers and repeating coils for communication purposes.

W. HERRIOTT was engaged in astronomical research at the Allegheny Observatory from 1914 to 1917. Research in astronomical and aerial photography at the Research Laboratories of the Eastman Kodak Company followed from 1917 to 1920. Between 1920 and 1925 he

was engaged in the development of military instruments and of optical apparatus for microscopy, photography and motion pictures at the Bausch and Lomb Optical Company. During the following three years he was in charge of the Scientific Department of the Fairchild Aerial Camera Corporation. In 1928 he joined the engineering department of Electrical Research Products, Inc., coming to the Bell Telephone Laboratories in 1929 to work on optical and photographic problems associated with sound picture apparatus development. In October 1936 he transferred to the Materials Group of the Electromechanical Division of the Telephone Apparatus Development Department.

A. H. Inglis, B.A., Yale University, 1914. Western Electric Company, Engineering Department, 1914–17. Signal Corps, A.E.F., 1917–19. American Telephone and Telegraph Company, Department of Development and Research, 1919–34; Bell Telephone Laboratories, 1934–. Mr. Inglis has been concerned with both equipment and transmission matters of station apparatus, latterly as Station Instrumentalities Engineer.

W. C. Jones, B.S. in Electrical Engineering, Colorado College, 1913. Western Electric Company, Engineering Department, 1913–25; Bell Telephone Laboratories, 1925–. As Transmission Instruments Director, Mr. Jones is concerned with the development of telephone instruments and similar devices.

H. C. Montgomery, A.B., University of Southern California, 1929; M.A., Columbia University, 1933. Bell Telephone Laboratories, 1929–. Engaged at first in studies of hearing acuity and related problems in physiological acoustics, Mr. Montgomery has been occupied more recently with the study and analysis of speech.

A. E. Ruehle, B.S., University of Idaho, 1930. Bell Telephone Laboratories, 1930–. Mr. Ruehle's work has been chiefly concerned with applications of the methods of physical chemistry to chemical analysis.

G. H. Stevenson, B.Sc. in Engineering, University of Glasgow, Scotland, 1906; Instructor in Electrical Engineering, University of Glasgow, 1906–07. Messrs. Barr and Stroud, Glasgow, 1907–11. Western Electric Company, Engineering Department, 1911–24; Patent Department, 1924–25. Bell Telephone Laboratories, Patent Department, 1925–. Mr. Stevenson's work has to do with patent matters

in the fields of wave transmission networks and radio transmission systems.

M. E. Strieby, A.B., Colorado College, 1914; B.S., Harvard, 1916; B.S. in E.E., Massachusetts Institute of Technology, 1916; New York Telephone Company, Engineering Department, 1916–17; Captain, Signal Corps, U. S. Army, A. E. F., 1917–19. American Telephone and Telegraph Company, Department of Development and Research, 1919–29; Bell Telephone Laboratories, 1929–. Mr. Strieby has been associated with various phases of transmission work, more particularly with the development of long toll circuits. At the present time, in his capacity as High Frequency Transmission Engineer, he directs studies of new and improved methods of carrier frequency transmission over existing or new facilities.

The Bell System Technical Journal

Vol. XVII *October, 1938* *No. 4*

Ultra-Short-Wave Transmission and Atmospheric Irregularities

BY C. R. ENGLUND, A. B. CRAWFORD AND W. W. MUMFORD

Results of an ultra-short-wave fading study are here reported. Transmission was carried out in the range of 1.6 to 5.0 meters, over a 70 mile (112.6 kilometer) ocean path, on 106 days during a period of two years. Both horizontal and vertical polarizations were used and during part of the time a 6-megacycle amplitude, 120-cycle, frequency modulated transmission was added, for the cathode-ray tube observation of the frequency characteristics of the radio path. On 45 mornings records were taken, on vertically polarized radiations, during the flight period of the Mitchel Field Weather Bureau plane.

Fading was found present practically all of the time. Amplitude changes up to 40 db and fading rates up to 5 fades per minute were found. Simultaneous transmission of the same wave in two polarizations, and of two waves of different wave-length in the same polarization showed that the horizontally polarized component was practically always, and the shorter wave-length one was usually the worse fader of the pair. The greater part of the time there was no correlation between the fading of these radiation pairs; occasionally, however, and for the slow, smooth amplitude, undulating type of fading, coincidence was observed. The frequency sweep patterns showed multiple signal components to be present, with various degrees of relative phase retardation.

A tentative explanation is proposed for these phenomena. This theory assumes the presence of a refracted-diffracted signal component, transmitted along the earth's surface and calculable in the manner of Wwedensky, Van der Pol and Gray, and one or more signal components reflected from air mass boundaries. The airplane results are shown to be in reasonable agreement with the frequency sweep observations. Boundary heights from 5.5 kilometers down to 1.9 kilometers are measured; below 1.9 kilometers other boundaries are indicated. The receiver band, flat over two megacycles, sets the low height limit of resolution of reflecting boundaries at 1.9 kilometers. Most of the boundaries are at the lower heights.

A discussion is given of some observations of signal fading at various wave-lengths which have been reported by other observers, and which are apparently referable to the same mechanism as is here proposed.

Introduction

IN an earlier paper [1] experimental data were presented which indicated that the transmission of ultra-short-wave signals was dependent upon the state of the atmosphere, in particular upon its water vapor content. The present paper contains the results of a continuation of this work where a two-year survey of ultra-short-wave transmission over a 70-mile (112.6 km.) ocean path was carried out. Transmission was had on 106 days during this period.

In planning this work, preparation was made for seeking a correlation between atmospheric structure and signal intensity; but from the very first transmission fading was found, and this fading was so persistent and intense that the work became essentially a fading study.

In the following paragraphs there are discussed, in the order named, Antennas and Locations; Apparatus and Operation; General Characteristics of Fading, with samples of records taken; Polarization Effect on Fading, with sample records; Wave-length Effect on Fading, also with records; Distance and Antenna Height Effects on Fading; Frequency Sweep Patterns of Fading, with sample records; and the logs taken during the flights of the U. S. Weather Bureau airplane for taking free air data. The presentation of experimental data is then interrupted to present a theory which explains several of the experimental observations. This is followed by further experimental results and checks, and concluding remarks.

Antennas and Locations

Figure 1 shows the layout of the radio circuit. The transmitter was erected at Highlands, New Jersey on the edge of a steep hillside. This edge made an angle of about 45° with the transmitter-receiver direction. Below the edge of the hill lay a strip of land slightly above sea level (seven to eight feet) and beyond was Sandy Hook Bay. The altitude at the antenna foot was 119 feet. The antennas consisted of a vertical rhombic terminated in its surge impedance with carbon lamps, a horizontal rhombic with the same termination, an unterminated inverted "Vee" antenna and a half-wave doublet. This doublet was equipped with a flexible transmission line which permitted it to be raised to the top of the antenna supporting mast. These antennas were supported on a central 60-foot (18.3 meter) lattice mast surrounded by four 30-foot (9.15 meter) poles.

The receiver was located on a plot of land at East Moriches, Long Island, New York. This plot was immediately at the edge of Moriches Bay and was only slightly (approximately four feet) above sea level. The same antenna equipment was supplied here as at the transmitter. Except for the transits across Sandy Hook, Fire Island Beach and Smith Point, the wave path was over sea water. A second receiving site at West Sayville, at the edge of Great South Bay, was briefly occupied, using portable receiving equipment. This site was 52¾ miles (85 km.) from Highlands.

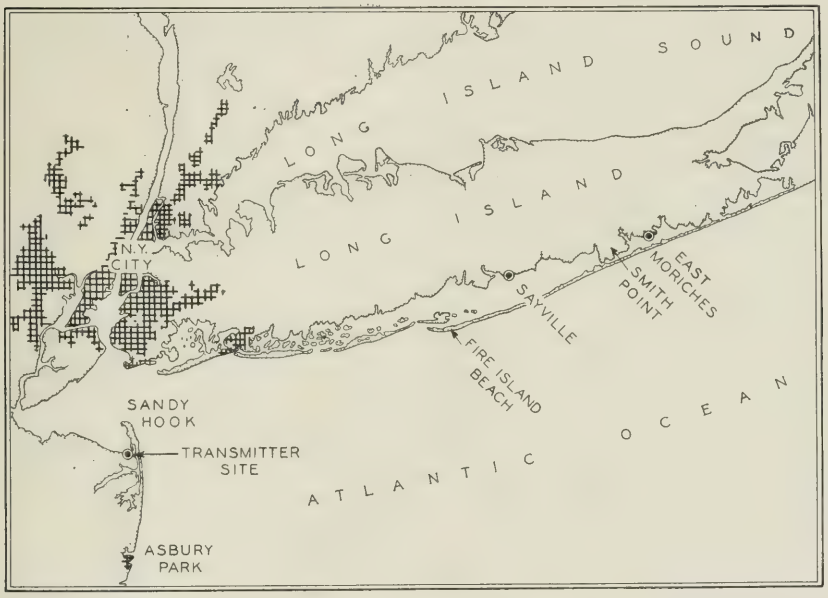


Fig. 1—Map of ultra-short-wave transmission path between Highlands, New Jersey, and East Moriches, Long Island.

Apparatus and Operation

In all, three transmitters were installed at Highlands. The first one, of 100 watts output, covered the wave-length range of 5.0 to 3.5 meters. It was equipped with a motor-driven single-turn short-circuit loop which, coupled with the tank circuit coil, produced a 120-cycle frequency modulation of six megacycles amplitude. For calibration purposes there was added a low-gain double-detection receiver which used an intermediate frequency of one megacycle and was connected so as to pick up an input from the transmitter. The beating oscillator of the receiver was set for the center of the transmitter frequency sweep and the receiver output triggered a gas tube connected

to the transmitter tube grids. The transmitter grids thus received a voltage pulse each time that the transmitter frequency passed through one megacycle above or below the beating oscillator frequency. Each transmitter frequency sweep was thus marked with two pulses spaced two megacycles apart.

The second transmitter had Lecher wire tuning elements, covered the wave-length range of 3.5 to 1.2 meters and had a power output of 30 watts at 1.5 meters. It was in operation simultaneously with the first transmitter for six months and then was replaced by transmitter No. 3.

The third transmitter was coil tuned, covered the wave-length range of 4.9 to 2.8 meters and had a power output over this range of 55 watts down to 35 watts. It was operated simultaneously with the first transmitter except for the first six months.

All three transmitters were arranged for voice modulation through a simple grid input, and the first one was thus used for one-way communication during the entire period of operation.

Normally, unmodulated waves were transmitted and were observed as rectified direct current in the output of the double detection receivers. These receivers had attenuators, variable in steps of 1 db, in the intermediate frequency amplifier circuits and the attenuators were geared to the pens of manual recorders. The operators kept the output current constant by means of the attenuators just mentioned, and there resulted a record of signal amplitude versus time. Some use was made of the Esterline-Angus type of milliampere recorder for automatic recording but no linear scale recorder of this type could handle the amplitude range of the fading encountered.

For the reception of the frequency modulated transmission a tuned radio-frequency receiver, with a three-megacycle band-width centered on 66 megacycles (4.55 meters), was constructed and its rectified output was applied to one pair of plates of a cathode ray oscillograph. A linear sweep voltage, manually synchronized with the transmitter 60-cycle power voltage, was applied to the second pair of plates. The oscillograph pattern thus pictured the frequency-amplitude characteristic of the radio circuit in toto. Over the frequency range where the receiver band was flat (two megacycles) the curve gave the apparent ether characteristic. With a motion picture camera this characteristic was permanently recorded.

Fading Characteristics, General

The fading was always slow compared with that observed on short waves. Except for the rapid fluctuations produced by airplane reflec-

tions, a record speed of $\frac{5}{8}$ inch (1.6 cm.) per minute was sufficient. This was our standard speed. Amplitude changes up to 40 db and fading rates up to 5 fades per minute were observed.

It is difficult to describe the fading in any other way than by the records. From a transmission standpoint a curve giving the per cent of time during which the signal is above the abscissa value is useful.

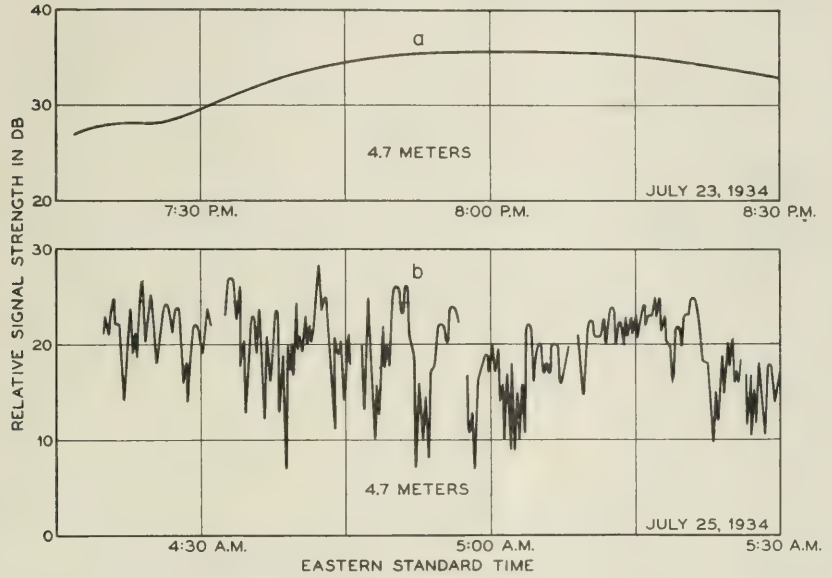


Fig. 2—Fading extremes, vertically polarized transmission; inverted "V" antennas.

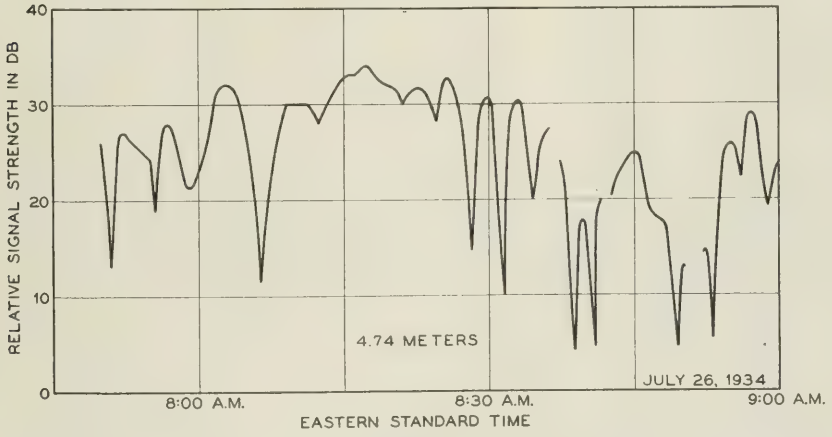


Fig. 3—Extreme amplitude, normal fading rate, vertically polarized transmission; inverted "V" antennas.

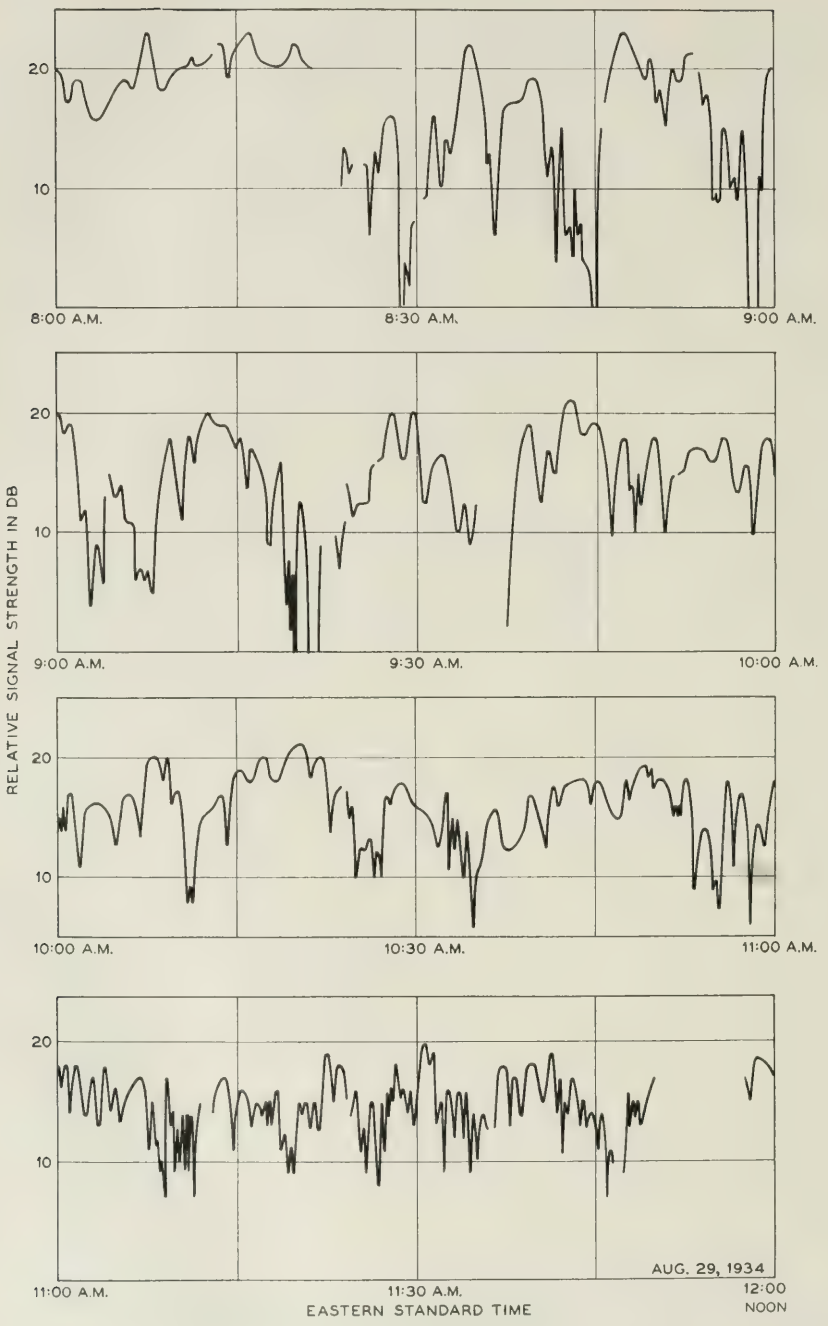
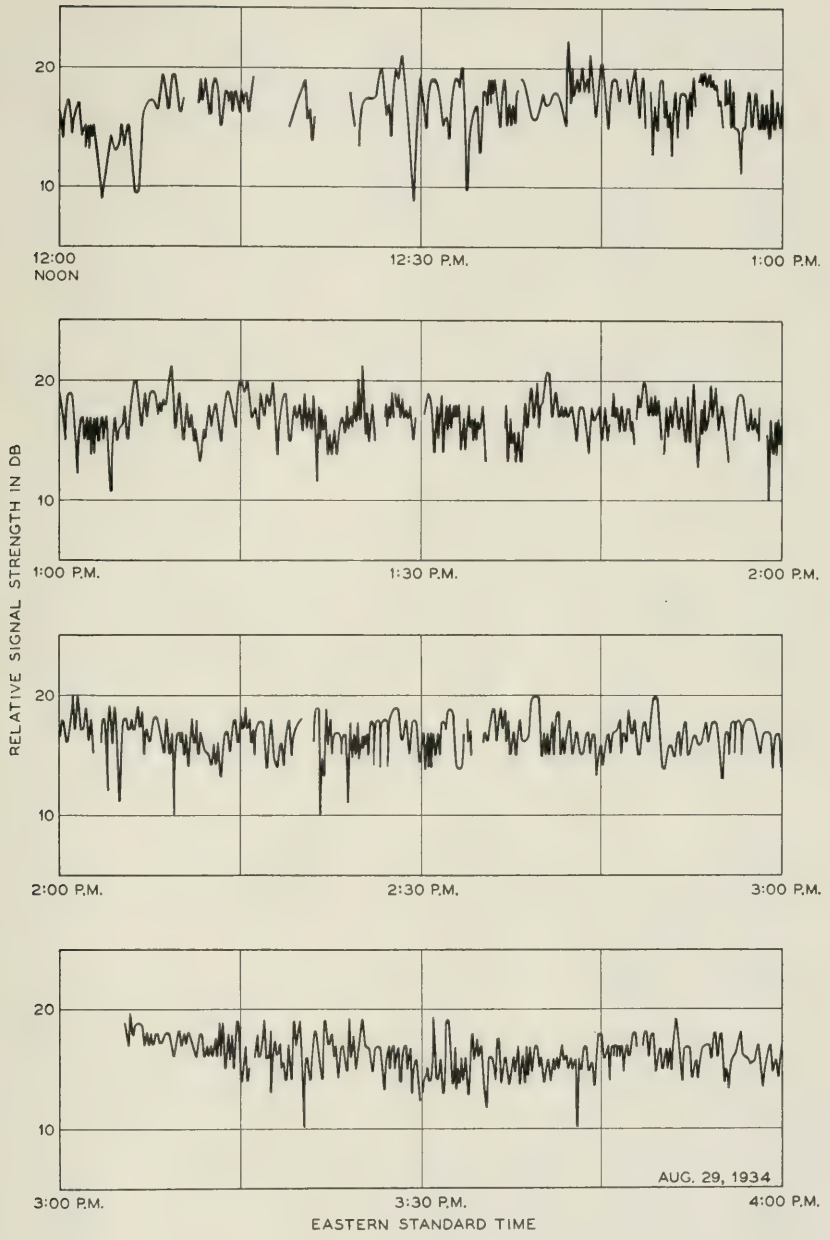


Fig. 4—Development of "scintillation" fading on vertically polarized



transmission, 4.74 meters wave-length; inverted "V" antennas.

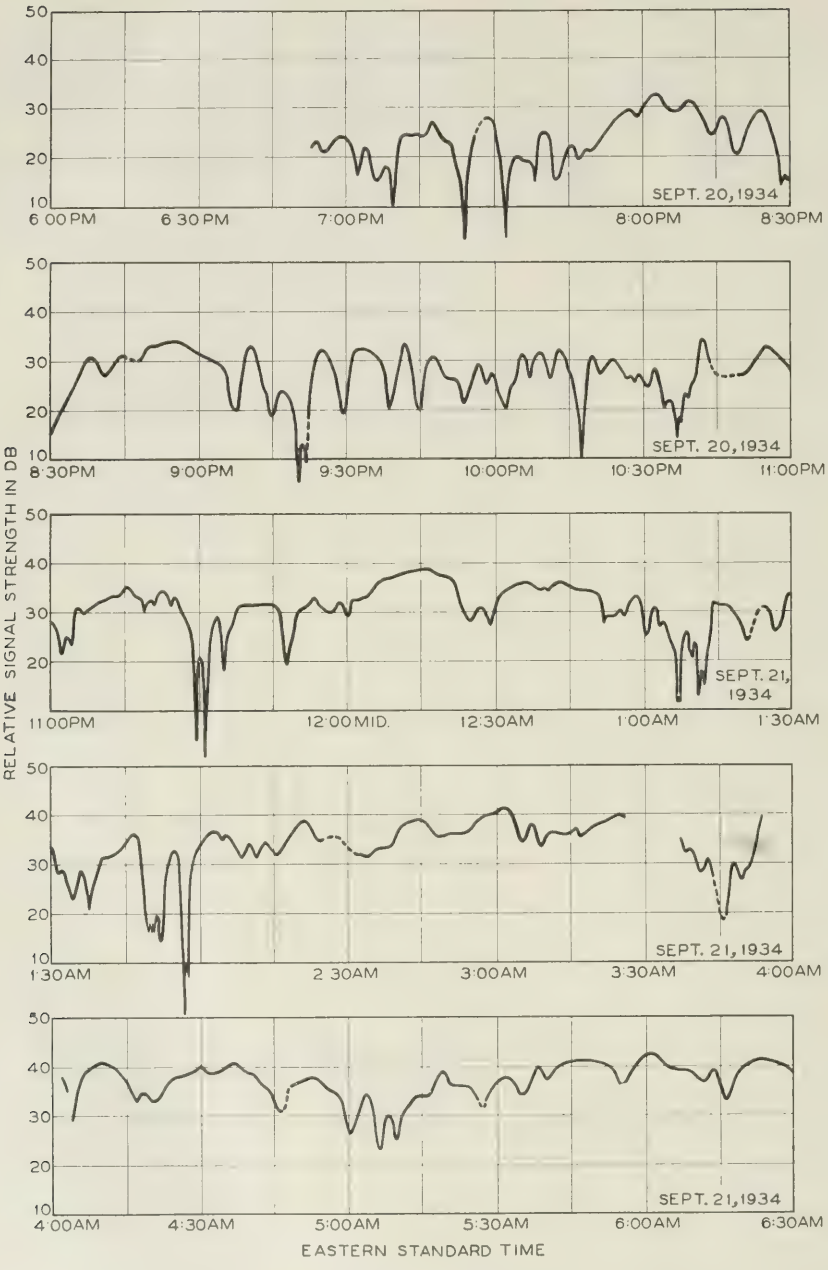
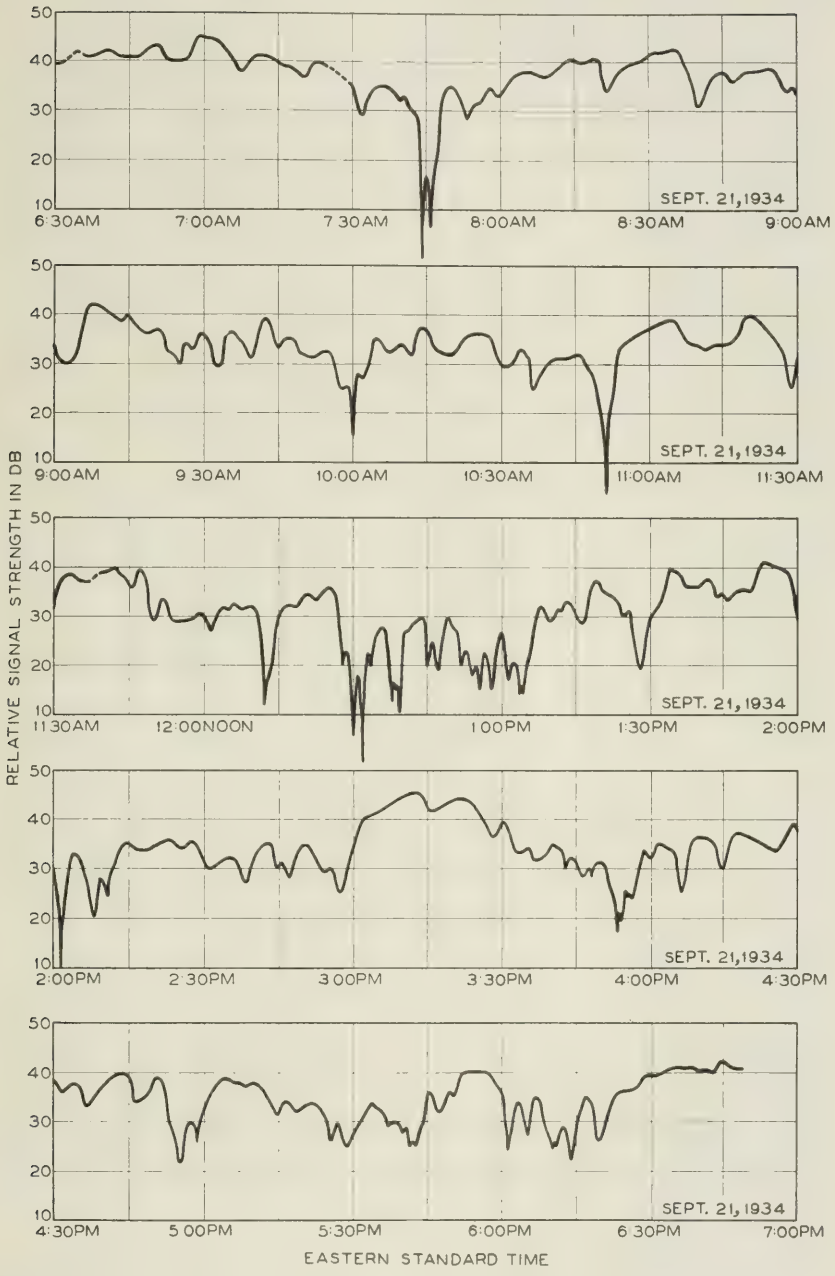


Fig. 5—Twenty-four hour run, vertically polarized



transmission, 4.74 meters wave-length.

Such a curve can also serve to check on the theoretical explanation of the cause of fading in certain cases. Thus if the fading is due to the combination of two radiation components in assigned random amplitude relation and arbitrary or random phase relation, a curve can be calculated from probability considerations and compared with the experimental curve.[2] Such a simple mechanism was inadequate for our fading most of the time. Moreover, the fading changed enormously from day to day. It is hoped that the samples given in the figures will give an adequate idea of this phenomenon.

Only rarely was fading practically absent for periods of an hour or two. Such a period is illustrated in Curve *a*, Fig. 2. Two days later the extreme fading of Curve *b* was recorded. It is significant, as will later appear, that the non-fading situation was the one of higher signal. The amplitude range of curve *b* is nearly normal; the fading rate is much greater than normal, for vertical polarization. In Fig. 3 the fading rate is normal but the amplitude range is excessive. In Fig. 4 a characteristic type of fading, which we have termed "scintillation," is recorded. In this case the fading, initially erratic and of a fairly wide amplitude range, subsides in a characteristic manner to a steady, fast rate oscillation, or scintillation, of moderate amplitude.

In Fig. 5 a 24-hour run is recorded. The rambling erratic character of the fading is well shown here. Characteristic deep short-period minima occur at intervals, occasionally they are twinned, some of them have a fine structure at the bottom. There are several "dropouts" where the signal practically disappeared.

No sunrise-sunset variations in fading were noticed, though looked for. Diurnal variations could not be established since automatic recording was not available. A seasonal falling off in average signal was noticed in the winter; the 1.6 meter wave, because of its normally low level, dropped below the noise level in the winter of '34–'35. No effect of ocean waves, clouds, or other visible weather phenomena could be established. It is true, however, that to be certain of the non-effect of such phenomena as clouds, a cloud observer at the midway point should have been present. In so far as cloud layers make air mass boundaries visible they may well affect the transmission. Cloud bottoms which represent merely the adiabatic dew point level should apparently not cause much signal reflection at these wavelengths.

Effect of Polarization on Fading

After some preliminary experimenting it was found that comparisons of two transmissions were worthless unless made on simultaneous recordings. The recorders were therefore fitted with telechron motors

operating on a circuit of the Patchogue division of the Long Island 60-cycle power network. The resulting timing was faultless and by transmitting the same radiation on crossed antennas, and receiving the vertical and horizontal components separately, a comparison was obtained.

In general the horizontal component showed the worse fading, more fades per minute and greater amplitude range. This was always true when the fading on vertical polarization was bad. There was then no noticeable coincidence between the two. When the fading had a smooth long period fade, or "roller," superposed on a short period oscillation, or "fine structure," there was at times coincidence between the roller components. Occasionally, with fine structure absent and

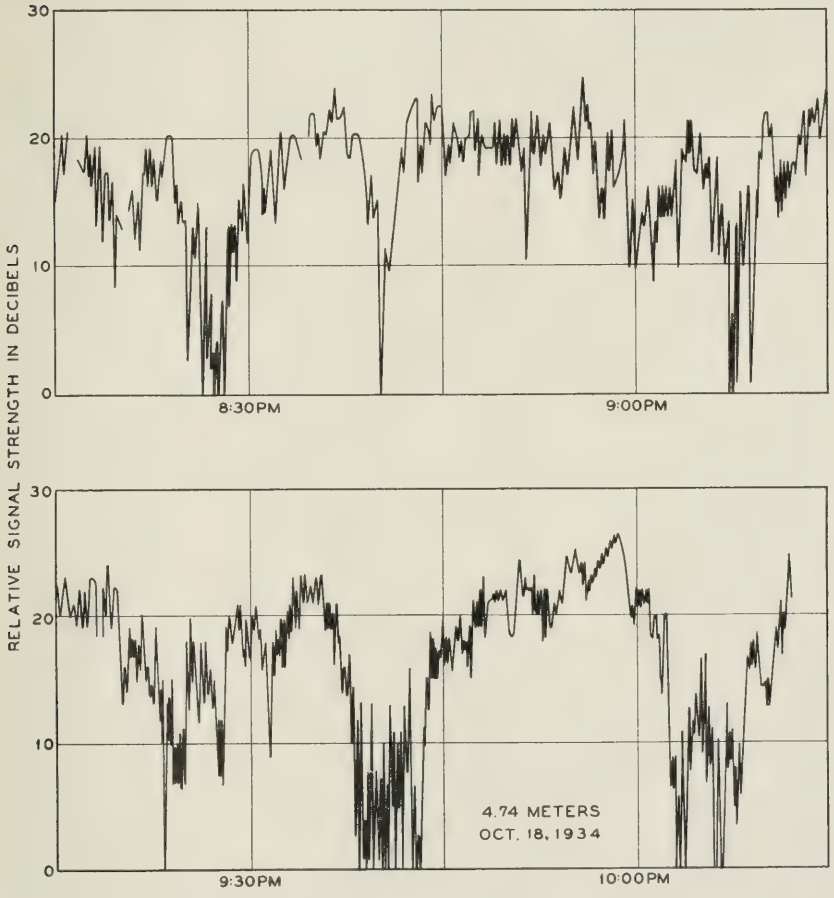


Fig. 6—Composite bad fading, horizontally polarized transmission.

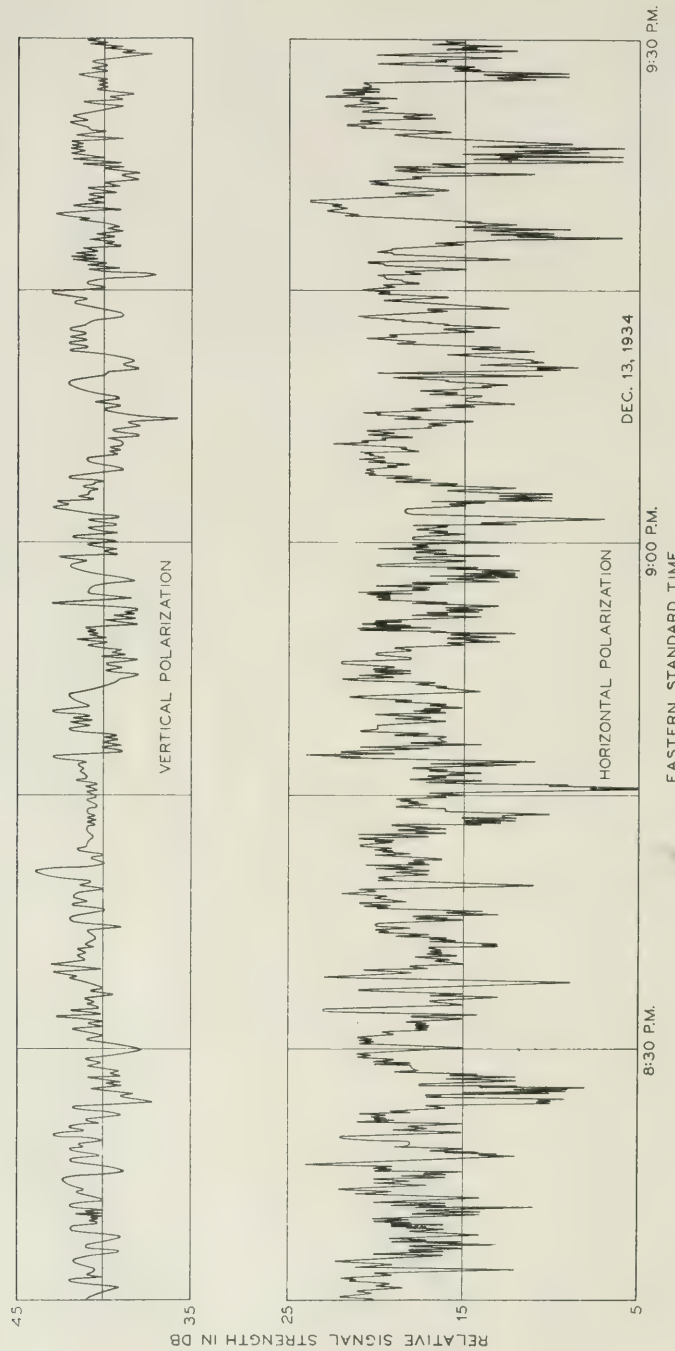


Fig. 7—Comparison of simultaneous bad fading on horizontally and vertically polarized transmissions, 4.74 meters wave-length.

moderate roller fading, a good coincidence between the two records resulted. This is discussed later.

Figure 6 is a sample of fading on horizontal polarization, at its worst. This particular specimen shows the superposition of roller and fine structure fading very well. No vertical-polarization record was taken along with this. Figure 7 shows a typical example of fading simultaneously observed on vertical and horizontal polarization during bad fading conditions. There is no coincidence. Figure 8, on the other hand, records an unusual condition when a mild roller type of fading shows a good coincidence on two polarizations.

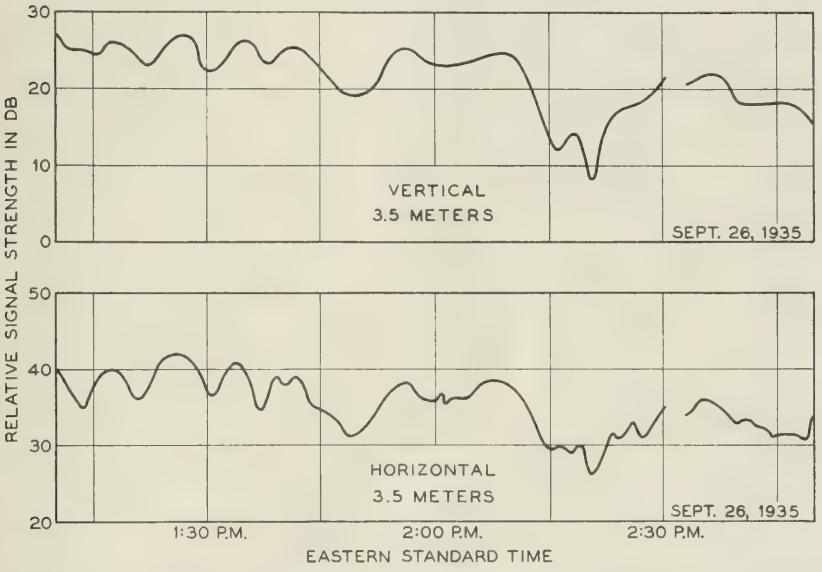


Fig. 8—Comparison of simultaneous mild roller fading on horizontally and vertically polarized transmissions.

Effect of Wave-length on Fading

The double wave-length records are not as contrasty as the double polarization ones. In general the shorter wave has the worse fading, either as higher fading rate, greater amplitude oscillation or both, and the greater the wave spacing the more certain this is to be true. Exceptions have occurred, however, where the fading was much the same, and one record was obtained where the fading rate on 4.7 meters was noticeably greater than on 4.5 meters.

Our first simultaneous records were taken at a wave-length ratio of 3 to 1 (4.7 to 1.58 meters) where the fading on the shorter wave was

always worse. The remaining observations were confined to wave-length ratios of 1.5 to 1 and less. A comprehensive set of records was obtained for moderate to small wave-length spacings, down to 1 per cent difference. These records are all for vertical polarization. The few records taken on horizontal polarization happened to be obtained when the horizontal fading was much worse than the vertical fading and the records are too rough for good comparisons.

For these small wave-length-difference records the types of fading are more likely than not to be similar on the two wave-lengths. That is, the fading rate and amplitude excursion will average up much the same. More rarely, there will be a similarity between the two fading tracks which is evident to the eye, sometimes as a "retarded" simi-

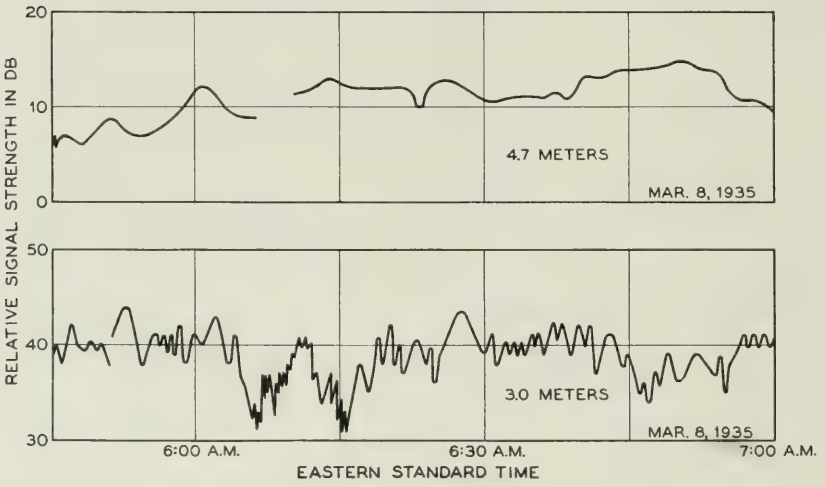


Fig. 9—Comparison of simultaneous fading on two well spaced wave-lengths, vertically polarized transmission.

larity. Occasionally, and usually on the roller type of fading, there will be a marked coincidence between the two records; this coincidence will be better the milder the fading and the smaller the wave-length spacing. Genuine identity was never recorded on different wave-lengths even down to 1 per cent difference. With scintillation, coincidence was difficult to demonstrate; a similarity on the major swings was all that was shown.

Figure 9 shows a very marked difference between 4.7 and 3.0 meter fading. This is one of our most contrasty records. Figure 10 shows very slow fading, on two occasions, with wave-length differences of

approximately 1 and 4 per cent respectively. There is good coincidence. Figure 11 shows active fading on short rollers for 4.7 and 4.65 meters, a wave-length difference of approximately 1.1 per cent. There is agreement in major features. Figure 12 shows a case of scintillation

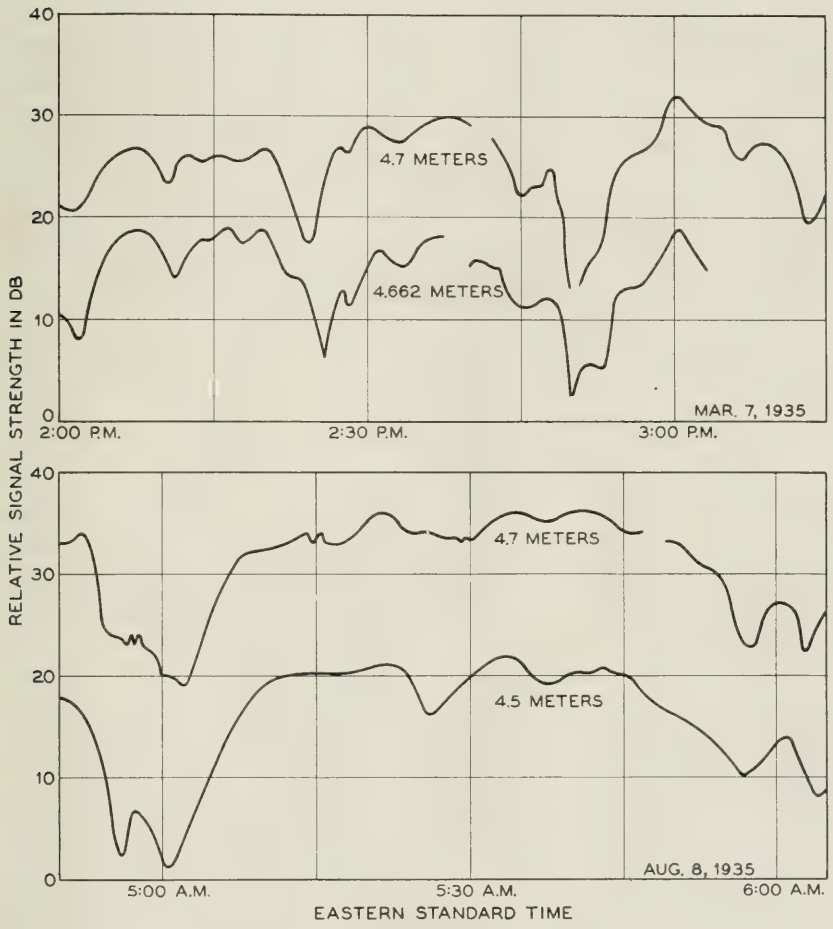


Fig. 10—Comparison of simultaneous slow fading on two slightly different wavelengths, vertically polarized transmission.

superposed on mild rollers, again for 4.7 and 4.65 meters. The time scale is here magnified three times. An in and out similarity can be seen, especially for the rollers. In the section on theory these similarities are further discussed.

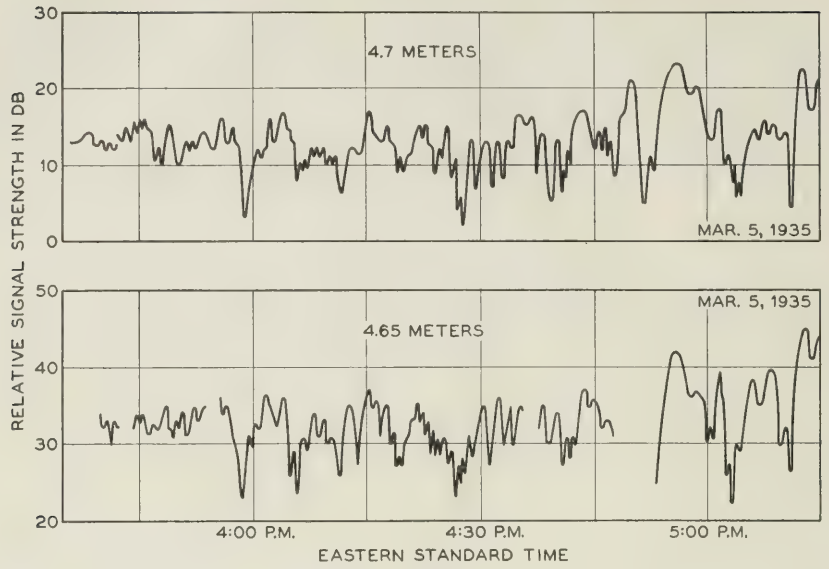


Fig. 11—Comparison of simultaneous active fading on two slightly different wave-lengths, vertically polarized transmission.

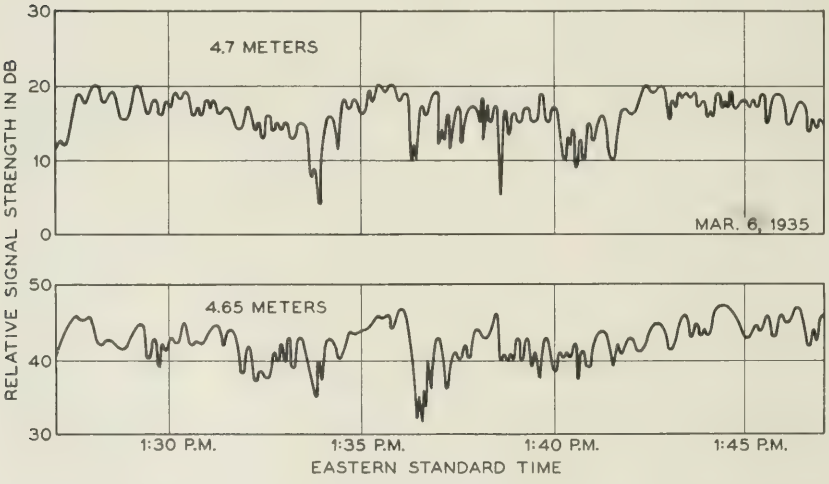


Fig. 12—Comparison of simultaneous "scintillation" fading on two slightly different wave-lengths, vertically polarized transmission. The time scale has been expanded.

Effect of Distance and Antenna Height on Fading

In planning this work a survey for a receiving site was made by means of a portable receiver in a car. Later, simultaneous reception

was had at East Moriches and West Sayville, on three days. The survey data were not sufficient to establish any proposition beyond the statement that the signal strength fell rapidly with distance, with the intensity of fading coming up as the signal fell. The simultaneous two-distance recording showed random fading between the two records with less fading amplitude at the shorter distance. The fading rate was about the same. Unfortunately the recording took place under scintillation conditions, thus giving very poor records for comparison purposes.

By mounting two linear doublets on the 60-foot lattice mast simultaneous recording at two heights was carried out. For the two doublets (horizontal, at 14 and 52 feet respectively), a signal level difference of 12 db was observed, in favor of the higher antenna. The fading on the two records was identical. It may be added that, on calibrating the car receiver at East Moriches before moving to West Sayville, identical fading records were obtained with the two antenna systems 150 feet (45.7 meters) apart and substantially broadside to the radiation.

Frequency Sweep Patterns of Fading

The frequency sweep patterns were of many types, from slow to fast fading and from shallow to deep fading. Apparent path differences from 600 meters down to a few meters occurred. The patterns were usually complicated, indicating that more than two components were present. There is no reason to believe, however, that they were not all due to wave interference.[13]

On three days the predominant pattern was simple enough to be referable to two waves with a path difference consistently greater than 75 meters. These will be referred to later. In Figure 13 are given three sample runs illustrating a two-component pattern, a three-component pattern with two of the components forming a small path difference pair, and a multiple component pattern. The receiver characteristic is dotted in on one curve of each set.

Logs During Weather Bureau Airplane Flights

On forty-five mornings recording was carried out during the period of flight of the Mitchel Field Weather Bureau plane. This plane takes off about dawn every morning, when flying is possible, and by means of a meteorograph obtains records of air pressure, temperature and humidity, up to an altitude of about five kilometers. A record of the fading, on 4.7 meters and vertical polarization, was obtained for each of these mornings. In addition, on twenty-six mornings frequency

sweep patterns were photographed at or shortly after the time of flight. These sweep patterns were all on horizontal polarization.

From the meteorograph data, kindly furnished us by the United States Weather Bureau, the dielectric constant of the air has been calculated [1] and plotted as a function of the altitude. On twenty-four days there were, above an altitude of 400 meters, changes in the dielectric constant curves equivalent to discontinuities of $\Delta\epsilon \geqq 10^{-5}$. Heights up to 3200 meters were recorded for these. Typical curves

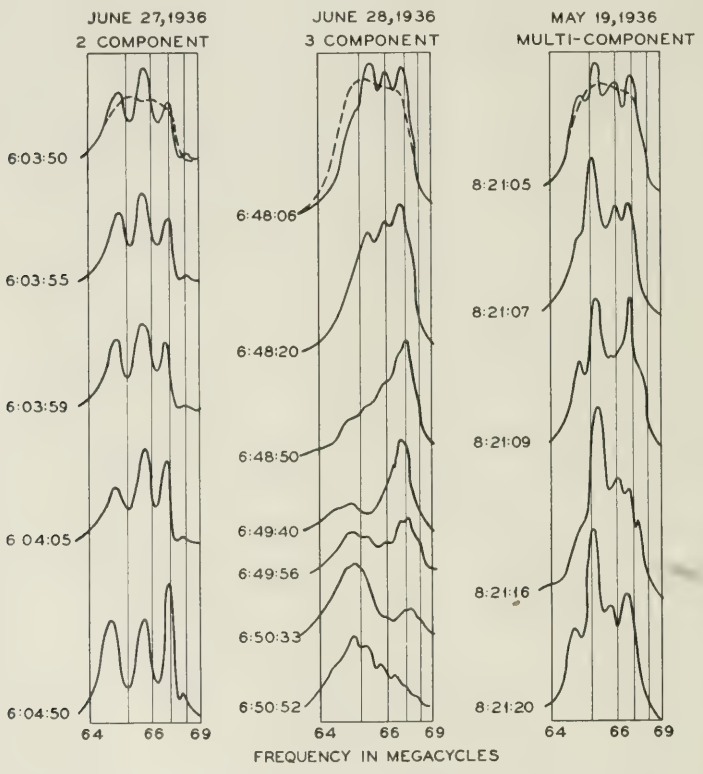


Fig. 13—Three sequences of frequency sweep patterns. Horizontally polarized transmission, 4.55 meters mean wave-length.

are given in Fig. 18 on the left-hand side. On four days there were small boundaries with $\Delta\epsilon < 10^{-5}$; on two days there were possible but not definite boundaries, the experimental points being too widely separated in altitude for precision; on five days there were possible boundaries below 400 meters altitude and on ten days the refractive

index-height relation was an approximate exponential one without any evident boundaries. These data will be referred to later.

Theory

The fading phenomenon was explicable in several ways. In our previously cited work [1] we found that variable atmospheric refraction was present, the airplane carried receiver being up where the refracted-diffracted field strength was high and dominant. In general variable refraction would be expected to be a slow phenomenon, operating in hours, or even days, rather than in minutes, and much too slow to explain five-cycle-per-minute oscillations, for example.

Another explanation was air-mass boundary reflection (or refraction),[3] such a boundary readily explaining the rate of signal variation. No Kennelley-Heaviside layer reflection was in question; this had been quickly ruled out by the experimental data. When, therefore, we elected to transmit the frequency modulated signal, already described, and the oscillograph revealed a cyclic maximum-minimum frequency characteristic of the other path itself, it was evident that there was no possibility other than wave interference left—interference presumably between a direct-diffracted and one or more boundary-reflected components.

These boundaries have apparently not been positively identified at longer wave-lengths and for that reason we have tried to get some further experimental contact with them. Attempts, since the closing down of the Atlantic Highlands-East Moriches circuit, to demonstrate an air-mass boundary, any boundary whatever, by high-angle transmission, have failed. No reflected components have appeared. Of course an illy defined, or diffuse, boundary will operate in this manner since only for near grazing incidence can such a boundary give the appearance of a discontinuity for the incident radiation.

If we assume such a boundary a few kilometers up, and assign to it a relatively small discontinuity in index of refraction, compared with that of an earth or sea water boundary, then the four components of Fig. 14 will be the only important boundary reflected ones for a radio circuit such as ours. We now, fortunately, have theoretical formulæ [4, 5] for computing the diffraction of an ultra-short-wave radiation around the earth and the amplitude of the direct-diffracted component can be calculated at once.

That is, it can be calculated at once if the air mass has no refractive bending effect upon the radiation trajectory. Since such a bending effect is certainly present at times, and is equally certainly variable, even if only slowly, it must be taken into account.

If the refractive index of the air varies as a power of the distance to the earth's center, it has been shown [6] that the actual state of affairs can be duplicated by a homogeneous atmosphere over an earth, the radius of curvature of which is greater than that of the actual earth and is calculable from the exponent of the height variation function. With this "effective" earth radius, the formulæ already mentioned become usable. If the air refractive index does not vary as a power of the distance to the center of the earth we must take that exponent which gives the best first order fit over the height covering the refracted wave front, the alternative being a prohibitive complication of the theory.

A plausible physical picture of the fading mechanism can now be set up. If we lump the four boundary reflected components in one, and plot as a function of the distance, we have curves "*A*" of Fig. 15.

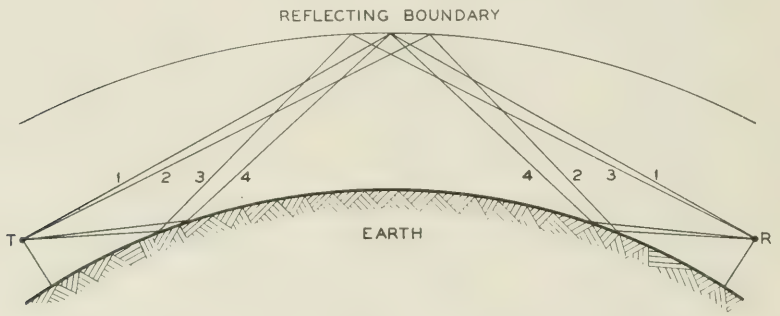


Fig. 14—Drawing illustrating the four components of a single reflection at an air boundary.

Curves "*B*" are the Wwedensky [4] * and Gray [5] theories. These are for our Highlands-East Moriches circuit with the average effective earth radius of 8500 kilometers and a 1500-meter boundary height. If we now imagine a receiver moving away from the transmitter we shall first traverse the zone of high "*B*" amplitude with no fading present. The signal amplitude will, for any given near-by point, and for any given antenna ampere-meters, depend on the height of the antenna above the ground and the ground constants. As the distance to the transmitter increases, the falling "*B*" curve approaches the rising "*A*" curve in ordinate and we enter a disturbed region where, for any instability of the boundary, more or less complete interference can result and fading will occur. (One such instability occurs when

* There is an error in the formula, as given by Wwedensky. It is corrected here. See appendix II.

a boundary with an irregular surface is carried past the reflection zone by the normal motion of the atmosphere.) A further increase in transmitter distance and the "*B*" or residual curve drops out of the picture leaving only the "*A*" curve and, presumably, fairly steady signal amplitude conditions. The location of these zones of undisturbed and disturbed signal will vary from day to day as: (1) the reflection coefficient and height of the layer change, (2) the effective radius of the earth changes. The effect of the height of the layer is shown in Fig. 16.

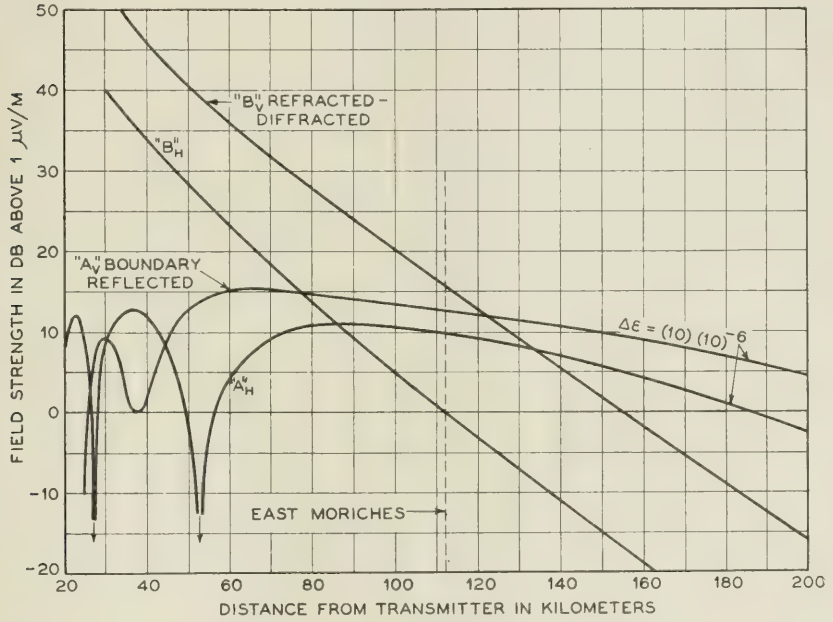


Fig. 15—Calculated curves for air boundary reflected and earth refracted-diffracted radiation components, in both vertical and horizontal polarization. Short doublet antennas, 1 kw. power radiated, wave-length 4.7 meters, $\sigma = 5 \times 10^{-11}$ (E.M.U.) and $\epsilon = 80$ for sea water. Height of transmitter antenna 42 meters, of receiver antenna 5 meters, air boundary height 1500 meters, effective radius of earth 8500 kilometers.

Since the major lobe of the polar characteristic of any simple antenna, such as ours, is directed forward and away from the earth, the signal intensity at the reflecting boundary surface will be comparatively high and will, in some measure, make up for a small reflection coefficient. For longer waves, such as broadcast waves, the high level of the "*B*" curve will move the disturbed zone so far out that the low residual signal level and the Kennelley-Heaviside layer reflections will conceal

or mask the atmospheric boundary reflections. Several observations which can be ascribed to such boundaries have nevertheless been published.[7, 8] Obviously, only boundaries lying considerably higher than those discussed here will give the path differences to produce the same type fading at these longer waves. At the same time the apparent diffuseness of a boundary will fall off with increase in wave-length, thus removing the restriction of reflection to near grazing incidence angles only.

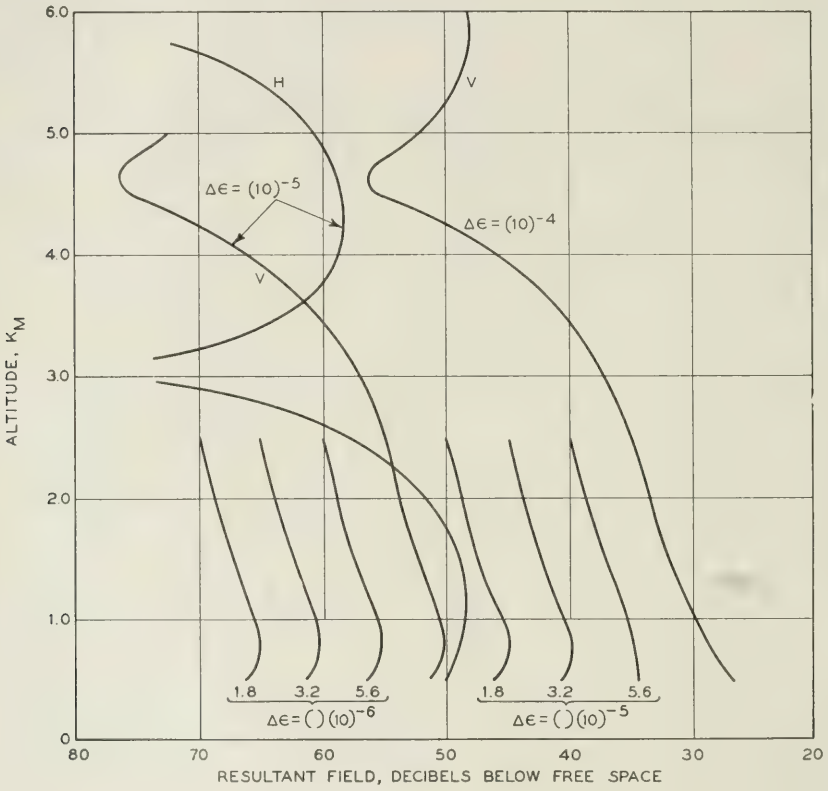


Fig. 16—Calculated field strength curves showing the effect of air boundary height and density on the reflected radiation component, for the Highlands East Moriches circuit. Transmission path 112 kilometers, over sea water, wave-length 4.7 meters, polarization both vertical and horizontal. Vertical antennas 42 and 5 meters high, horizontal antennas 45 and 9.5 meters high, respectively.

This tentative mechanism also explains several other observed features. Thus, for a given type of boundary instability, the fading rate will increase as the wave-length decreases. Furthermore, since the slope of the "B" curve increases as the wave-length decreases, the

disturbance zone is effectively moved nearer the transmitter and the probability of increase in fading amplitude is enhanced. The usual increase of fading with decrease in wave-length is thus explained. When the wave-length difference is small, on the other hand, the fading type should be much the same on both wave-lengths, as was generally found. The lack of coincidence would arise from the fact that the path difference being a considerable number of wave-lengths, a small wave-length change can introduce a marked randomness in fading without appreciably affecting the type.

As has been mentioned earlier, a multiple of reflecting boundaries is the normal condition, rather than that of a single boundary. This circumstance, without invalidating the explanations already given, makes a further elaboration of the theory possible. The "roller" type or component of fading, in particular, requires explanation. In addition to the smooth signal modulation, from which the name has been derived, this type of fading is characterized by showing more or less frequent deep minima or drop-outs and these are often distinctively twinned. Further, the roller component is that component of fading which shows coincidence, in spite of wave-length or polarization differences. Such coincidence indicates small path difference and this is what we have when a double boundary or stratum exists. Such a stratum would give two "*A*" components and, if of variable thickness, would, as it was carried along by the prevailing air currents, give the steep, deep, minima at phase opposition thickness. Further, if the stratum contour were that of a hump, thick enough to carry the second "*A*" component past phase opposition to the first one, the twinned minima would result as the hump entered and left the reflection zone. Occasionally the two "*A*" components would add properly, with the residual "*B*" component, to give complete extinction, a result less likely from the phase addition of a single "*A*" and the "*B*" component.

This explanation of "roller" fading assumes, tacitly, that the "*B*" component is, at the time, relatively subdued, that is, the disturbance zone has moved inwards due to an increase in the reflection coefficient of the layers or to a decreased "effective" earth radius. The fine structure often appearing at the bottom of a prolonged roller minimum corroborates this, the mutual cancellation of the two "*A*" components having uncovered, so to speak, the weaker "*B*" component with its much shorter traversed path.

With the "roller" condition characteristic of high "*A*" component signal amplitude, the "scintillation" condition would be characteristic of low "*A*" component signal amplitude, the relatively steady "*B*"

component having superposed on it a small amplitude, variable phase, "A" component. A relatively low mean amplitude value and the coincidence of scintillation conditions with conditions of convective instability of the atmosphere would thus be explained. All the scintillation records came on days of relatively high wind and convective instability. A turbulent atmospheric condition would dissipate or attenuate any boundaries, especially the lower ones. The rapid flutter about the mean amplitude value is the normal expectation from a high, turbulent, low reflection coefficient boundary.

Our two polarization results are qualitatively explicable on the mechanism proposed. As can be seen in Fig. 15, the change from vertical to horizontal polarization results in a relative lowering of the "B" curve without much change in the "A" curve, which should result in increased fading. For our circuit and a boundary at 1500 meters the relative "B" vs. "A" drop is 13 db.

As Fig. 16 shows, the variations of the "A" components with height are markedly different for the two polarizations. The "A_V" component falls steadily with height up to 4700 meters; the "A_H" component has a deep and sharp minimum at 3000 meters after which it rises again. Since most of our observations concerned boundaries at 2000 meters or less, this high altitude disparity between "A_H" and "A_V" does not affect our explanation. The disparity between vertical and horizontal fading should be much more marked for high boundaries than for low boundaries.

Further Experimental Curves and Checks

The curves given have illustrated the variability in the fading, a variability which no short period of recording can encompass. The tentative explanations proposed have been shown to be in accord with several of the features characteristic of this fading. Certain other experimental results will now be adduced which offer further verification along somewhat different lines.

For the forty-five mornings on which simultaneous recording was carried out during the United States Weather Bureau plane flight, we have calculated, from the airplane data, the values of the "A" and "B" components. As stated earlier, there were twenty-four days when boundaries above 400 meters altitude, and of sufficient distinctness to be fairly accurately estimated ($\Delta\epsilon \geqq 10^{-5}$) were shown by the meteorograph records. For these the "A" components have been computed. By taking the dielectric constant gradient for the first half kilometer, the effective earth's radius was determined and inserted in the Wwedensky formula to give the "B" component. These calculated values

("*A*" component as triangles, "*B*" component as circles) are plotted on Fig. 17 together with the maximum and median * observed values. These latter are joined by lines. For the 10 mornings on which no boundaries were evident the calculated "*B*" component appears to be some 8 db higher than the observed values. With this correction the agreement between observed and total calculated fields is fairly good. A partial explanation of this 8 db discrepancy may lie in the fact that the ocean water trajectory assumed in the calculation differs from the actual one by the land terminals and the three tongues of land intervening.

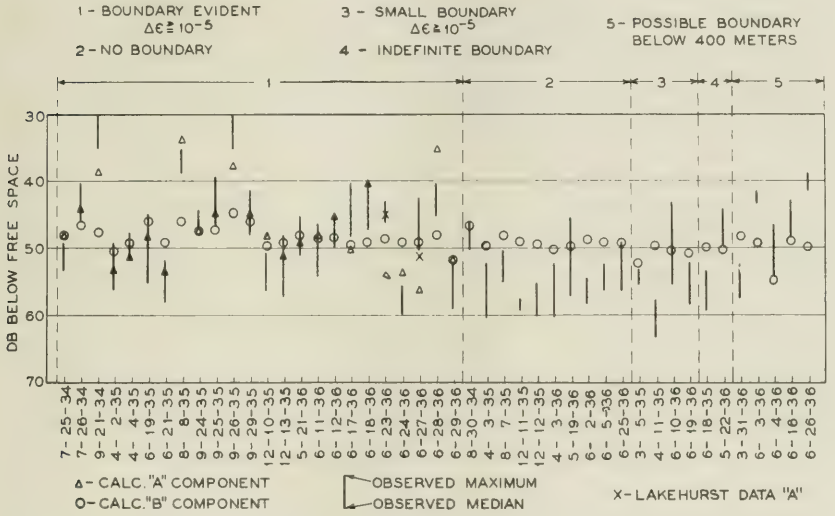


Fig. 17—Comparison of "*A*" and "*B*" radiation components, calculated from the U. S. Weather Bureau free air data, with measured maximum and median signal strengths. Vertically polarized transmission.

On the twenty-six morning frequency sweep runs there were only three on which the predominant sweep pattern was simple enough to be interpreted as due to two components with path difference greater than 75 meters. For those days a series of measurements of the film patterns was made by determining the frequency spacing between a maximum and a minimum and calculating the resulting path difference and boundary height. The dielectric constant-height function was also calculated from the Weather Bureau data. These curves are

* The signal is half of the time greater and half of the time less than its median value. For random phase with "*B*" component equal to "*A*" component the resultant median value signal is $\sqrt{2} \times A$ or 3 db up; it falls from this value to "*A*" as "*B*" decreases to zero.

plotted in Fig. 18 with the calculated boundary heights set down at the right hand, spread out in time of observation. The boundary height coincidence is pretty definitely located in this manner. Many of the more complicated frequency sweep patterns carried a fine struc-

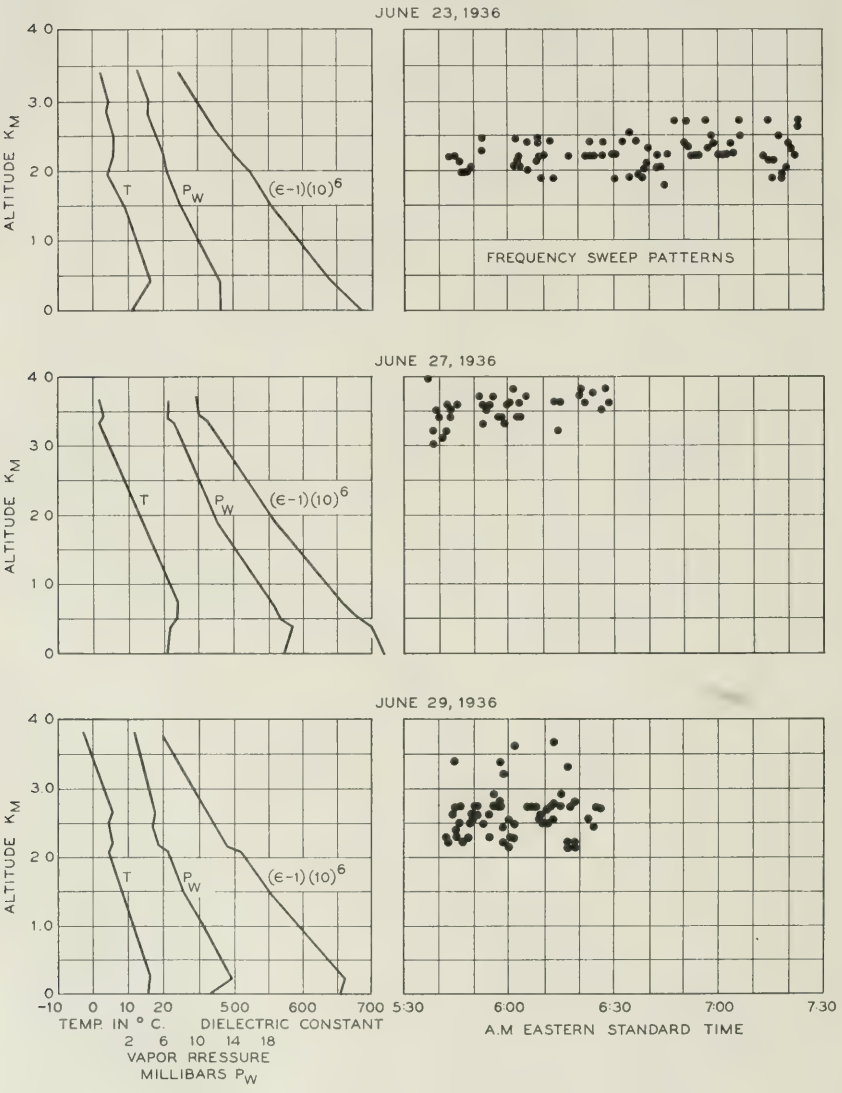


Fig. 18—Comparison of boundary heights shown by the U. S. Weather Bureau free air data, with boundary heights measured from frequency sweep patterns. Horizontally polarized transmission.

ture which indicated weak boundaries at higher altitudes up to, roughly, 5.5 km.; most of the patterns, however, were characteristic of layers below two kilometers. The path difference corresponding to two kilometers is 85 meters. The theoretical limit of resolution of the amplifier band for a maximum to minimum frequency spacing is $\Delta l = 2(C/\Delta f)$ where Δl = path difference, C = velocity of light and Δf = frequency band. For $\Delta f = 2 \times 10^6$ cycles this gives 75 meters, and hence boundary heights at and below 1900 meters are unresolvable by our receiver. It is a remarkable result that the bulk of the disturbing boundaries should lie so low.

It was mentioned earlier that several observations referable to air-mass boundaries have been published. In addition there have been reports, for three consecutive years, of long distance ultra-short-wave reception by American amateurs [9] during the month of May. We have copies of the U. S. Weather Bureau atmosphere cross-sections for several of these days and have been curious enough to examine them. On May 9, 1936, during the long distance amateur reception, there was an extensive boundary at 4 km. between an upper Superior air mass and a lower Tropical Gulf air mass. On May 15, 1937, a similar boundary at 4–5 km. had a Superior air mass above a wedge of Transitional Polar to Tropical Atlantic air. Below this at 3–4 km. lay a Transitional Polar Continental air mass.

On June 11, 1936, when Colwell and Friend [7] report an extra strong 0–2 km. "*C*" reflection, a subsiding Transitional Polar Pacific air mass lay above a Transitional Polar Continental air mass with the boundary at about 1.5 km. On June 29, 1936, when they reported a very strong 3.5 km. "*C*" reflection, there existed four wedge-shaped air masses with a Superior air mass over a Transitional Polar air mass at 3–4 kilometers. The wave-lengths used were 186, 125 and 86 meters approximately.

These coincidences may or may not be significant but it is very questionable that any boundaries at such altitudes are due to either electron or gas ion distributions.

The characteristic properties of North American air masses have been published,[11] as average summer and winter values, and show some marked seasonal differences. The greater dielectric constants for summer conditions are due chiefly to greater water content.

For a single air-mass distribution, horizontal stratifications are at a minimum and the radio transmission is via the "*B*" component. This component can be calculated from the corresponding effective earth radius. The table below gives this radius for three important air mass types.

Air Mass Type	Effective Earth Radius	
	Summer	Winter
Tropical Gulf—T_g	$1.53 \times R$	$1.43 \times R$
Polar Continental—P_c	$1.31 \times R$	$1.25 \times R$
Superior—S	$1.25 \times R$	$1.25 \times R$

"R" = actual earth radius

The boundaries between different air-mass types furnish discontinuities adequate for radio reflections. The greater the stability of the boundary, the more abrupt it is likely to be. In general, when "S" air overlays either "T_g" or "P_c" air, the resulting boundary is stable. Possible discontinuities, for the three types discussed, may be summarized in the following table. Here the positive sign means that the radiation originates in the more refractive medium. For stability the lower medium is the denser though not necessarily the more refractive.

Altitude	$\Delta\epsilon \times 10^6$					
	Summer			Winter		
	S/T_g	S/P_c	T_g/P_c	S/T_g	S/P_c	T_g/P_c
1.0 Km	100	20	−80	55	25	−30
2.0 Km	50	10	−40	50	15	−35
3.0 Km	30	10	−20	35	10	−25

Concluding Remarks

The characteristics of this seventy-mile circuit indicate that for ultra-short-wave transmission it rates as a long distance one. If we assume that the air refraction is on the average such that the effective earth's radius is 4/3 the actual one, then the receiving station lay 1400 feet below the line of sight from the transmitter. This is equivalent to 0.57° below the horizon. The reception, using high effective-height antennas, was good; there was, however, very little lee-way left, above set noise, for reception with simple doublet antennas. Any longer circuit will require to be terminated on elevations such as to keep the intermediate horizon height down. The fading was too slow to be noticeable on amplitude modulated speech unless a deep minimum or drop-out occurred.

The circuit was probably unusable for television, most of the time. A system adhering to the R.M.A. standard [12] of 441 lines on an inter-

laced 60-cycle scanning will have a unit time element of 0.17 microseconds. This corresponds to a path difference of 51 meters and only a fraction of this is necessary to produce a ghost. A rough estimate of the boundary height range involved in our fading is one-half to five-and-one-half kilometers. The corresponding path difference range is 8 to 580 meters. As the fading records show, no matter whether the "A" or the "B" component predominated, the other component was usually present in amplitude only second to the other. It may be pointed out that where a standing wave system exists,[10] reflected components with much larger path differences than those recorded here are almost certain to be found.

Appendix I

In the Wwedensky [4] paper the author applies his theory to one of the experimental curves from a previous paper of ours. He uses the normal earth radius "R," however, without any correction for air refraction. If we assume, as a more probable effective earth radius, the value $4/3R$,[6] the agreement with our curve is markedly improved.

Appendix II

In the first Wwedensky paper, *Tech. Phys.* U. S. S. R. Vol. 2, p. 632, 1935 eq. (7, 1) the sign of the term $|\tau_m|^2 \sin 2\theta_m$ should be minus.

Appendix III

The fading produced by moving bodies such as airplanes has been referred to in one of our earlier papers.[10] It happened one day, during the present investigation, that fading of this type appeared when mechanical recorders were being used and, by speeding up the paper, a record in two polarizations was obtained. The airplane itself (or other cause) was not visible. The results are given in Fig. 19. Again the horizontal component was the worse one. At first the two fadings, both fine and coarse components, were in step; later they passed entirely out of step where the fading was so rapid as to smear the paper. These "airplane" fadings were observed, off and on, at other times but were not recorded.

References

1. Englund, Crawford and Mumford, *Bell System Technical Journal*, Vol. 14, p. 369, 1935.
2. Brown and Leitch, *Proc. I. R. E.*, Vol. 25, p. 583, 1937; Norton, *Proc. I. R. E.*, Vol. 26, p. 115, 1938.
3. Ross Hull, *Q.S.T.*, Vol. 21, p. 16, 1937, May.
4. B. Wwedensky, *Tech. Phys.* U. S. S. R., Vol. 2, p. 624, 1935; Vol. 3, p. 915, 1936; Vol. 4, p. 579, 1937.

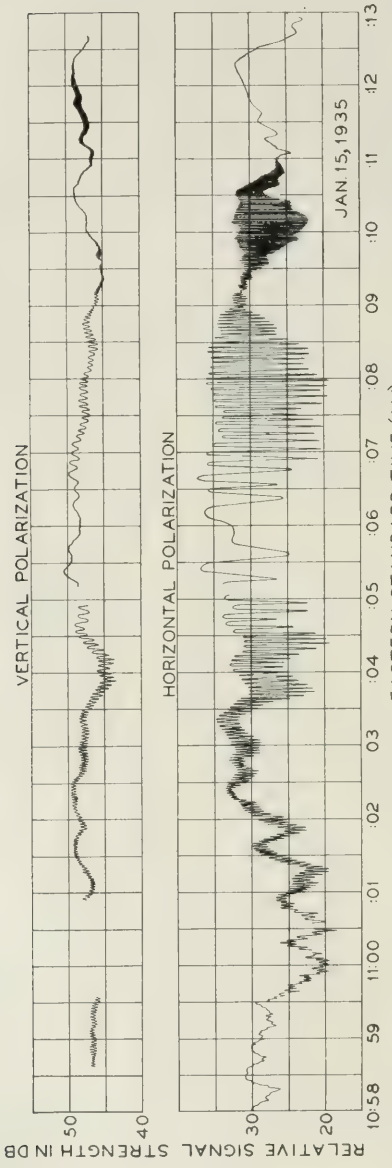


Fig. 19—Records, in two polarizations, of a transient high speed fading attributable to radio reflection from a moving airplane.

B. van der Pol and Bremmer, *Phil. Mag.*, Vol. 24, p. 141, 1937; Vol. 24, p. 825, 1937.
5. Miss M. C. Gray, paper to be published.*
6. Schelleng, Burrows and Ferrell, *Proc. I. R. E.*, Vol. 21, p. 427, 1933.
7. Colwell and Friend, *Nature*, Vol. 137, p. 782, 1936; *Phys. Rev.*, Vol. 50, p. 632, 1936; Colwell, Friend, Hall and Hill, *Nature*, Vol. 138, p. 245, 1936; Friend and Colwell, *Proc. I. R. E.*, Vol. 25, p. 1531, 1937.
8. Watson Watt, Wilkins and Bowen, *Proc. Roy. Soc.*, A, Vol. 161, p. 181, 1937.
9. *Q.S.T.*, Vol. 21, p. 27, 1937, July.
10. Englund, Crawford and Mumford, *Proc. I. R. E.*, Vol. 21, p. 464, 1933.
11. H. C. Willett, *Bull. Amer. Meteor. Soc.*, Vol. 17, p. 213, 1936.
12. Beal, *Television*, Vol. 2, p. 15, 1937, R.C.A. Inst's. Press.
13. Englund, Crawford and Mumford, *Nature*, Vol. 137, p. 743, 1936.

* The case of vertical polarization is treated by references 4, that of horizontal polarization by reference 5.

Amplitude Range Control

By S. B. WRIGHT

The art of controlling the amplitude range of telephone signals involves recognition of certain characteristics in addition to those used to specify the performance of ordinary transducers. Fundamentally, three kinds of characteristics are necessary to distinguish different range control devices. They are (1) the steady-state input-output characteristics, (2) the time actions, and (3) the range over which they function. In some cases, several secondary characteristics may be of interest, but they need not be considered in determining to which class a particular device belongs.

This paper discusses and classifies these characteristics.

Introduction

IN a "non-linear" transducer, the output power is not proportional to the input power. Consequently, the ratio of maximum to minimum power at the output differs from that at the input. But the ratio of maximum to minimum power is an expression of amplitude range. A device designed to alter this ratio may be called a *range controller*.

In telephony the term *range controller* includes many devices [1] having specific names, such as limiters, volume control devices, range reducers, compressors, vogads, expandors, etc. These devices have many properties in common with telephone repeaters, and a repeater may be considered as a special case in which any non-linearity which may exist between the output and input is unintentional.

The purpose of one type of range controller is to reduce the range of significant intensities of signals applied to a telephone circuit so as to ease the requirements of the transmission medium with respect to overloading and noise interference. Such a device is placed at the transmitting end of the circuit. When the range is compressed at the sending end of the circuit it may sometimes be desirable to expand it at the receiving end to the original range. This is done with a device having, in general, the same dynamic characteristic as the compressing device, but a range change which is complementary. The purpose of the expandor is to reduce the noise heard by the listener as well as to compensate for whatever characteristic signal modification occurred in

[1] For numbered references, see end of text.

the process of compressing the original wave. Sometimes an expandor is used at the receiving end to reduce the gain in the intervals between the main signals even when no compressor is employed. This is an example of using a range controller to correct defects in the medium.

As is well known, the performance of a repeater is specified by such characteristics as impedance, amplification, frequency band, noise, and output carrying capacity. The performance of a non-linear device involves some additional characteristics. The primary ones are (1) the slope of the input-output curve, which tells how the range is changed, (2) the dynamic operation, which tells the manner in which the output varies with time following a given change in input, and (3) the range, which tells the region over which the device exercises control.

It may be helpful to imagine a range controller as an amplifier in tandem with an adjustable attenuator, the loss of which may be changed either instantly or slowly to follow in some predetermined fashion changes in the signal. For simultaneous operation, this device could put out a wave which is a simple function of the input, but if the operation were delayed by a definite interval the device would be required to respond in a complex fashion in accordance with a recollection of what had occurred in the signal during the delay period. Such delayed adjustment would be very crude for intervals comparable with the periods of fundamental speech frequencies. To obtain practical regulation of the delayed type it is necessary to increase the delay beyond this range and base the control upon the amplitudes of the syllables. When the delay is increased to a point where it is comparable with the syllabic periods its usefulness is again reduced.

Part 1—Control Ratio

Fundamental Characteristics

Figure 1 shows how waves may be altered by a device having a given output-input characteristic, assuming the operation is instantaneous. As this figure is plotted on a db scale, only the stronger portions of positive values of the wave are shown. A similar diagram could be drawn for negative values. The output-input characteristic, although a straight line in this kind of diagram, would of course be parabola-like if plotted on a current or voltage basis. By selecting points, such as A (or B), on the input wave and determining the relative outputs A'' (or B''), the corresponding resultant wave is obtained. In this case, the resultant has a flatter top than the original sine wave, and this illustrates the capabilities of the device in increasing weak signals with respect to the strong ones and also suggests that distortion

may accompany the transformation. Such effects depend upon the slope of the output-input characteristic.

The control ratio of a range controller might be defined as the output range in db divided by the input range in db within the non-linear region of interest. The ratio is obtained in such a way as to eliminate transient effects, i.e., using steady-state sine waves.

Typical Control Ratios

Figure 2 shows some typical output-input characteristics for various transducers having control ratios between zero and infinity. While

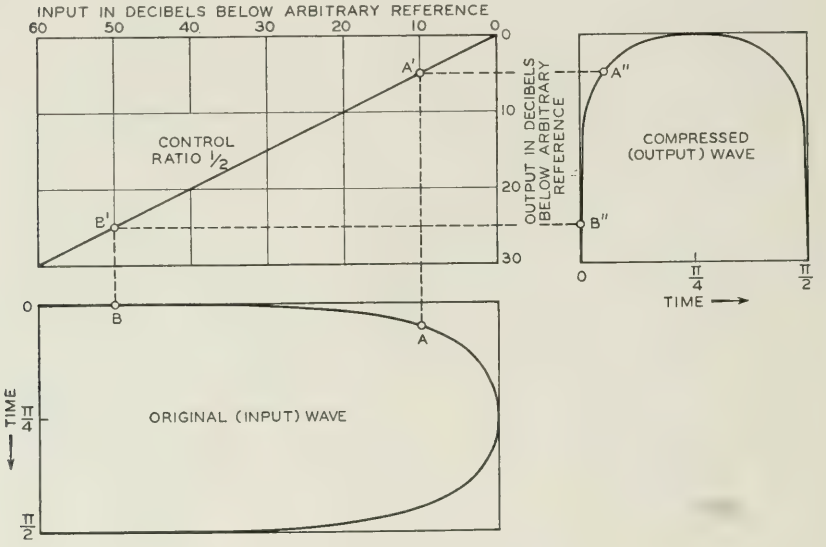


Fig. 1—The signal modification caused by a non-linear transducer depends upon the slope of the output-input characteristic.

these typical characteristics are straight lines there is nothing to prevent a range controller having a control ratio which varies with input. However, when complementary action is required at the receiving end it is more readily obtained when the control ratio is constant. Also, some physical elements used in the design of range controllers are most readily adapted to a straight line characteristic.

Compressors (that is, devices having control ratios less than 1) may be divided into two classes: (1) Complete * and (2) Incomplete. In a complete compressor (control ratio = 0) the output is held constant within the range of the device. This control ratio gives a maximum

* This is not usually of practical interest but is useful as an ideal limit of operation.

of possible noise improvement and also a maximum of signal modification. There is, however, no information in the compressed signal

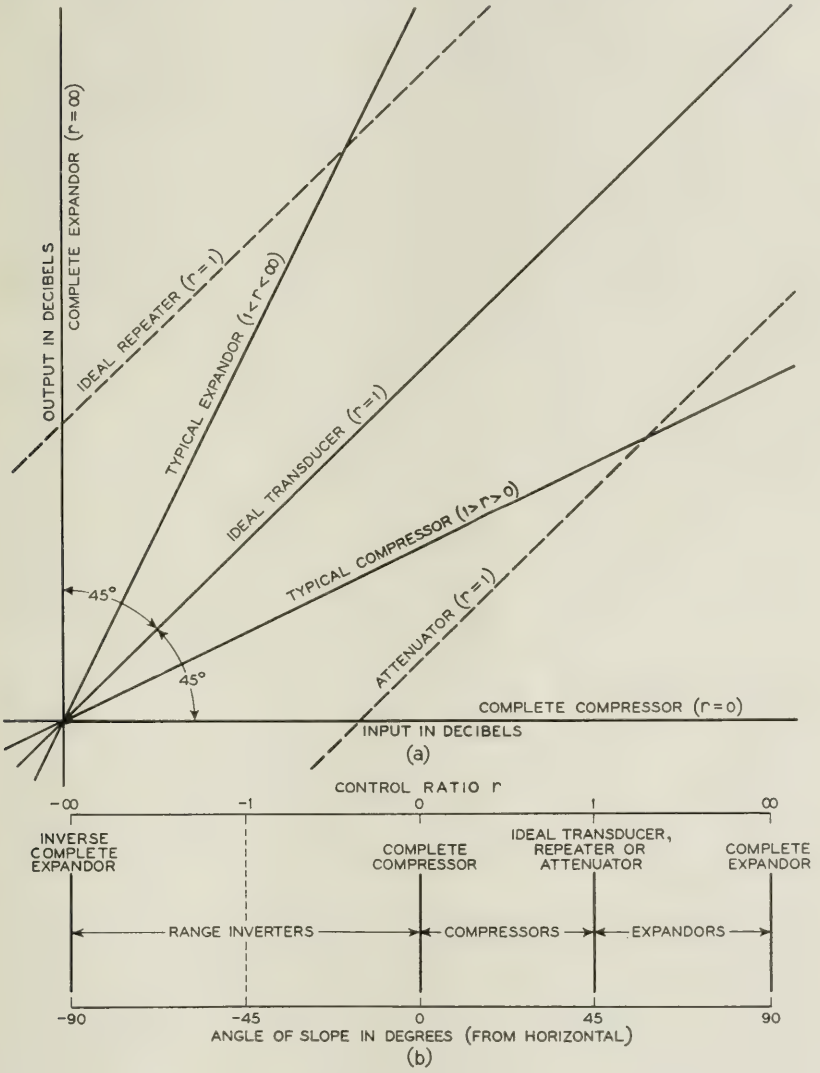


Fig. 2—If transducers are classified with respect to the slope of the input-output characteristics, several fields of action with definite demarcations result.

which would serve to indicate how much compression occurred. Consequently, if it were desired to restore the original range, it would be necessary to transmit this information in addition to the compressed

signal. The gain of the restoring device would be guided by this auxiliary information. Hence, the device used to pass the information along is called a "pilot channel." Various types of pilot channels are listed in Part 4 as secondary characteristics of the control.

When the control ratio is between 0 and 1 the compression is incomplete. A wave compressed in this manner has the property of being able to cause re-expansion at the receiving end since the output amplitude bears a definite relation to the original, assuming constant transmission over the intermediate circuit.

In the field of expandors having a control ratio between one and infinity the signal modification is opposite to that of compressors. Thus a convenient method is available for restoring the original wave shape by using an expandor having a control ratio which is the reciprocal of that of the compressor at the sending end.

Effects of Control Ratio

The control ratio is useful in determining the effectiveness of a device in improving transmission in the presence of noise in the medium. When noise alone is acting on the device, the noise determines the action in a manner similar to speech. When both noise and speech are present, the action is determined by the sum of the two. Thus, room noise applied with the speech will be compressed or expanded exactly as if it were part of the speech. In the case of a compressor used at the sending end of a noisy circuit, an input range of say 60 db might be compressed to 20 db, by using a control ratio of 1/3 over the entire input range. At a point where the strongest signals are unchanged, the weaker signals would then be 40 db stronger than when the compressor was omitted. The improvement of the signal and applied noise with respect to noise in the medium thus depends on the difference in ranges at the input and output which depends on the control ratio.

A large part of the usefulness of an expandor is in changing the apparent ratio of speech to the noise heard in the absence of speech, since the noise is generally weaker than speech and is made even less compared to speech by expansion. This is in spite of the fact that at any instant the signal-to-noise ratio is the same at the output as at the input. When the noise is comparable with the speech in amplitude, or when the noise is so weak as to be negligible without a controller, there can be no improvement in the noise conditions in using these devices. Between these two limits, the noise improvement rises to a maximum value also determined by the control ratio, and the time actions and range to be discussed.

A receiving range controller also changes variations in the transmission medium in proportion to the control ratio.

Part 2—Time Actions

Instantaneous Control

A device having a given control ratio might have its gain changed simultaneously with the applied e.m.f. The signal modification would become greater as the control ratio departed farther from unity and the modified signals would approach rectangular wave shapes at the limiting control ratios. Unless instantaneous compression is limited to a very small part of the signal range, an incomplete instantaneous expandor (inverse rooter) is required at the distant end which does the reverse of what is done at the transmitting end to restore the signal to substantially its original form. Due to the characteristics of the compressed signals, however, a transmission bandwidth without appreciable amplitude or phase distortion of two to three times the normal is necessary for high quality transmission.

Rectified Control

To avoid the necessity of transmitting such a wide band of frequencies, as well as to permit the use of a single device without restoring, in which case the distortion is limited to a value which is permissible from the standpoint of a listener, practical devices do not operate instantaneously. Instead, the gain is controlled by the charge on a condenser, which is controlled by rectified waves. The action of such an arrangement will now be discussed.

Consider a wave formed by subtracting two sine waves equal in amplitude, one having a frequency 10 per cent less than the other.* A portion of such a wave is shown in Fig. 3*a*. This wave is equivalent to a cosine wave of frequency one-half the sum of the two frequencies, as shown by the instantaneous voltages of Fig. 3*a*, multiplied by a secondary wave (envelope) of frequency one-half the difference of the two original frequencies.

The instantaneous voltages of the wave of Fig. 3*a* vary from a positive maximum through zero to a negative maximum. Curve *a* of Fig. 4 is a summation of most of the instantaneous e.m.f.'s of Fig. 3*a* with respect to their occurrence. About 99 per cent of the instantaneous voltages are in the ranges shown, the remainder being in the range between the upper and lower halves of Fig. 4.

* This illustration is not directly comparable with speech, but it contains some of the attributes which are comparable in this analysis, besides being readily reproducible and relatively simple.

Figure 3*b* indicates values for the same wave in which the negative ordinates have their signs reversed by means of an ideal full-wave rectifier. The resulting wave contains frequencies which were not present in the original, prominent among them being second and higher harmonics of the original. The range of instantaneous values shown

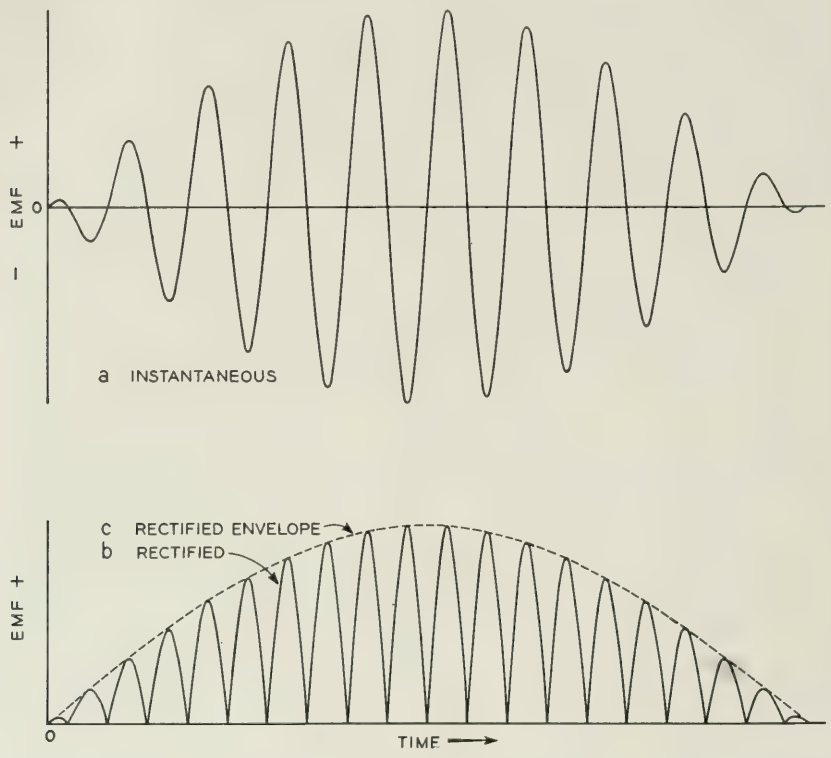


Fig. 3—A wave's amplitude varies from positive maximum to negative maximum. If symmetrical, the amplitude may be expressed as varying in only one direction from zero to maximum by rectification.

on Curve *b* of Fig. 4 is only half that of the instantaneous voltages. About 99 per cent of the values lie in a 60 db range.

The instantaneous values of the envelope of the rectified wave follow curves 3*c* and 4*c*. In speech the envelope is composed of many rather low frequencies which are determined by the rates of enunciation of syllables. For this reason they are sometimes called the syllabic frequencies. If it were possible to make the control vary as a function of the envelope, the result of using a control ratio of 1/2 on the wave of Fig. 3*a* would be as shown in Fig. 5*c*. This was

obtained by multiplying the original wave by a factor which is inversely proportional to the rectified envelope. For comparison, the original wave is shown in Fig. 5*a*, and the result of instantaneous compression

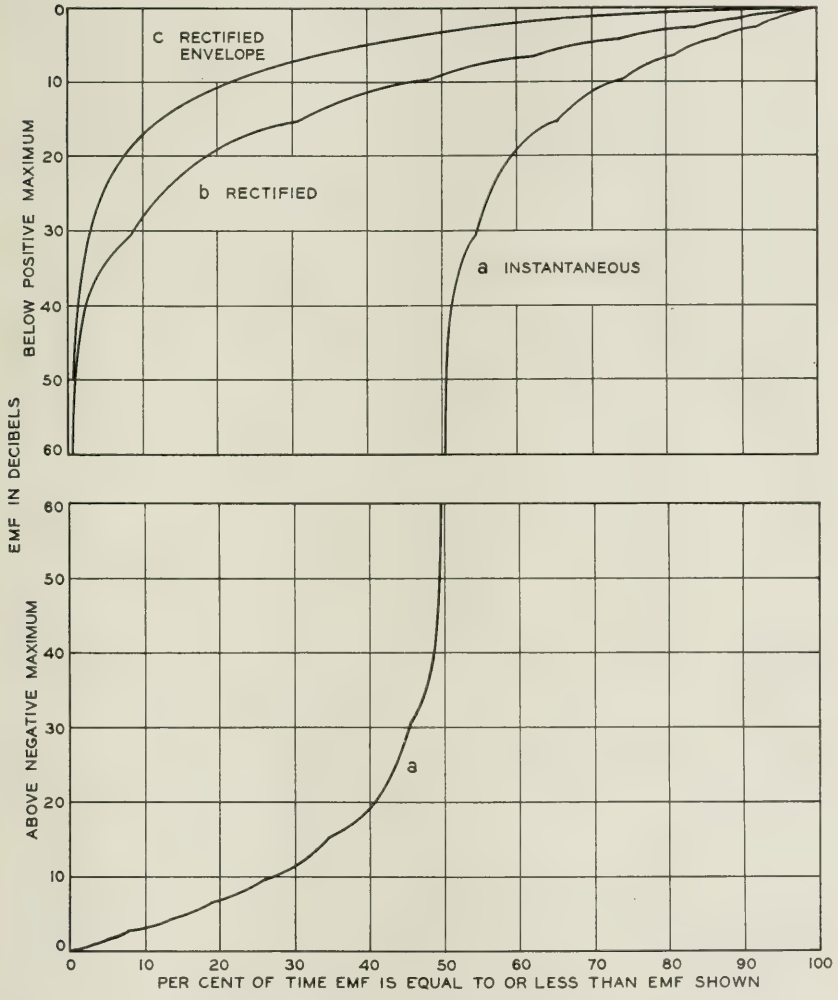


Fig. 4—The amplitude range of the wave of Fig. 3 is infinite on a db scale but most of the values are bunched in a much smaller range.

by the same control ratio in Fig. 5*b*. It is assumed that the arbitrary reference voltage which is not changed by compression corresponds to the maximum value of the input wave, although any other value might be used instead. It is evident from Fig. 5 that envelope com-

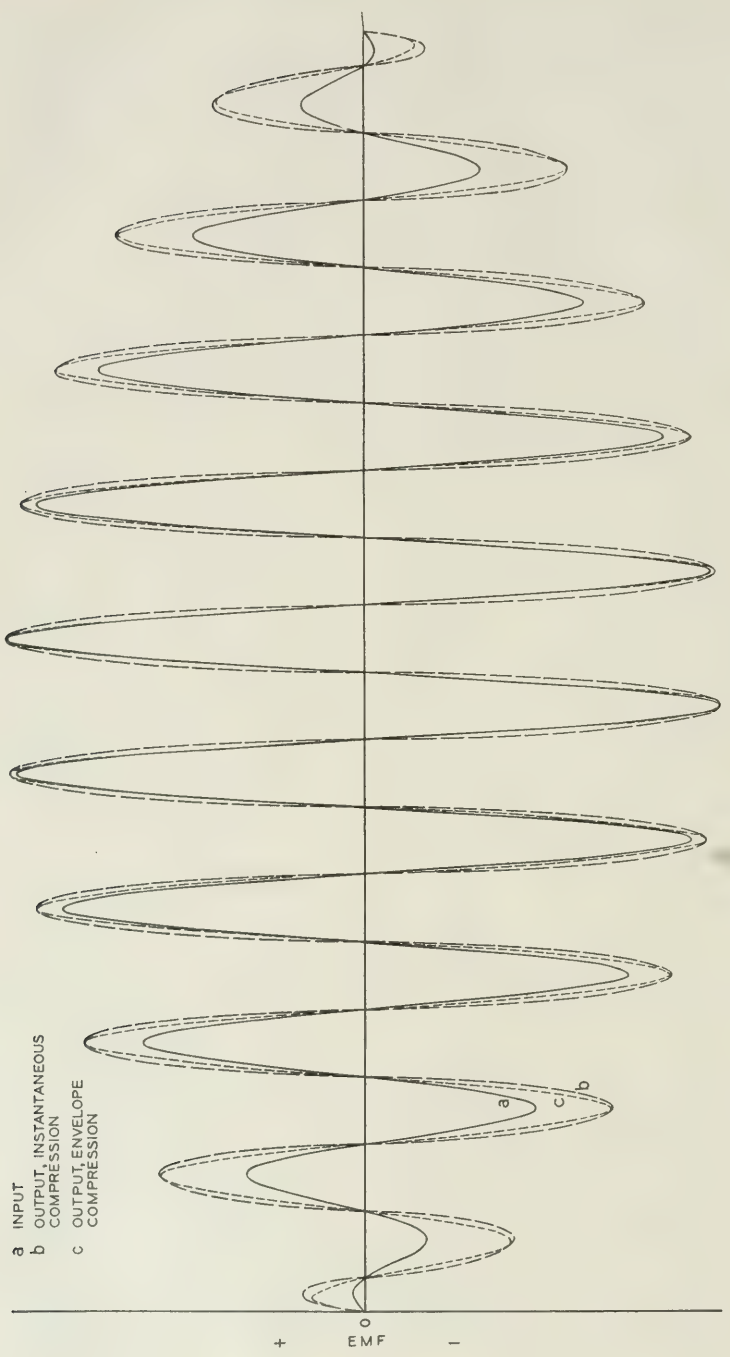


Fig. 5—A wave is compared with two corresponding compressed waves using a control ratio of 1 : 2. For instantaneous action the loops are relatively wider than the original, while in envelope compression the shape is more nearly retained. In both cases the amplitudes of the weak peaks are increased relative to the strong ones.

pression would result in less distortion than instantaneous compression. The extra frequencies formed that were not present in the original wave are the envelope frequencies, so that the additional band required to transmit this wave faithfully is negligible.

Dynamic Operation

The measurements [2] and adjustments of speech amplitudes in common use are made with devices that integrate the effects of the wave over certain time intervals. They do this in a rather complicated manner, however, so that it is difficult to express the resulting quantities in terms that are generally understood.

In the measuring instruments the rectified voltages are impressed on a condenser before being sent through a meter. The readings of the meter are, therefore, proportional to the voltage on the condenser modified by the damping of the meter. The voltage is made up of the sum of the effects of all the instantaneous voltages that have been applied to the condenser from the beginning of time to the instant under consideration. These effects die out so rapidly, however, that the instantaneous voltage on the condenser is practically determined by the voltages received in the immediate past. The condenser may be said to have a memory but a short one. In range control devices, the condenser forms the voltage which determines the amplification of the device.

To distinguish this voltage on the condenser from the applied voltage at any instant, we may call the former an "impression" of the original wave. If the time constant RC is small we get strong impressions similar to the rectified applied wave and its envelope, and if it is large we get weak impressions quite different from the applied wave but something like the rectified envelope.

Figure 6 shows the impressions of the wave of Fig. 3*a*, using four different values of time constant RC as compared to P, the period * of the envelope. Figure 7 shows smoothed summation curves of the impressions of Fig. 6 formed during the time $P/2$. Comparing this with Fig. 4, it is evident that the "bunching" effect for the distribution of impressions is largely between those for the rectified instantaneous and envelope curves. For the longer time constants, i.e., weak impressions, this is not the case for the weaker e.m.f.'s.

* This is twice the duration of Fig. 3, since only half a cycle is illustrated. It is assumed that C is completely discharged at the time this wave is applied. In practice, the rectifier impedance varies with the applied e.m.f. so that the results are not as simple as in this illustration. In general, the time actions are different depending on whether the applied e.m.f. is increasing or decreasing.

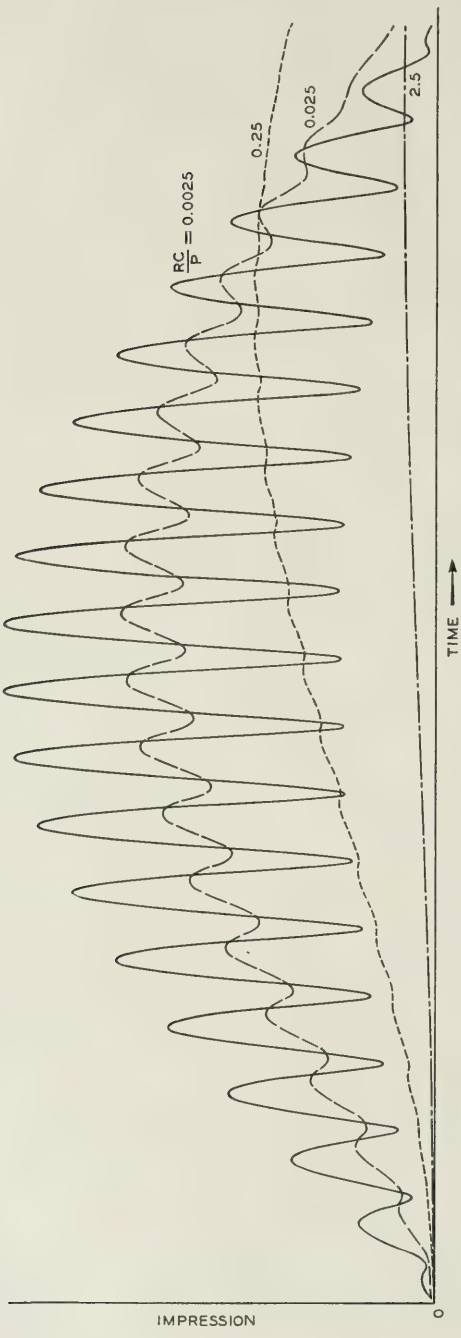


Fig. 6—An integration of Fig. 3*b* differs from the original wave and lags behind it by increasing amounts as the time constant is increased.

Referring again to Fig. 6, it will be seen that for the two smaller values of RC/P the impression curves are composed of (1) the envelope frequency, (2) double the fundamental frequency, and (3) a small delay which can often be neglected. An approximation to envelope compression is therefore possible by choosing RC/P to be in the proper range, i.e., .0025 to .025, and making the output vary as a root or power of the impressions thus formed.

Figure 8 shows the result of compressing the wave of Fig. 3*a* by using the impressions of Fig. 6 to determine the amplification. It was

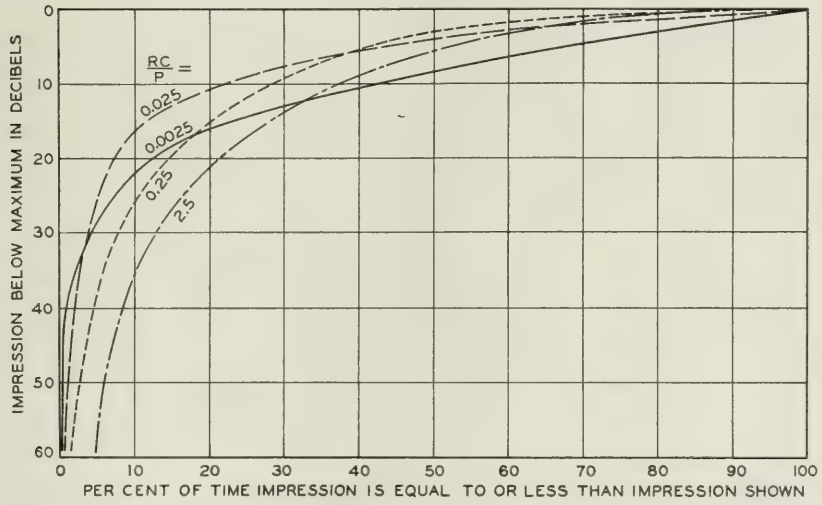


Fig. 7—The amplitude ranges of the impressions shown in Fig. 6 are bunched differently, depending on the time constant. A "volume" measurement means that a given impression is exceeded a small percentage of the time. In speech the peaks are relatively higher than in the wave illustrated.

assumed that the amplification varies in inverse proportion to the square root of the impression. The resulting waves for $RC/P = .0025$ and .025 (medium impressions) are recognizable as something like the original wave. However, for the larger values of RC/P (weak impressions), the distortion at the beginning of the wave is quite large. This is because the impressions are formed so slowly that a longer time is required to drive the gain down to the desired value.

In order to compare impression compression with instantaneous compression, the ordinates of Fig. 5 and 8 were plotted in Fig. 9. This shows that the greatest possibilities of bunching the waves into a narrow range result from the use of instantaneous compression (*b*), since the ratio between any value and the maximum is modified by the

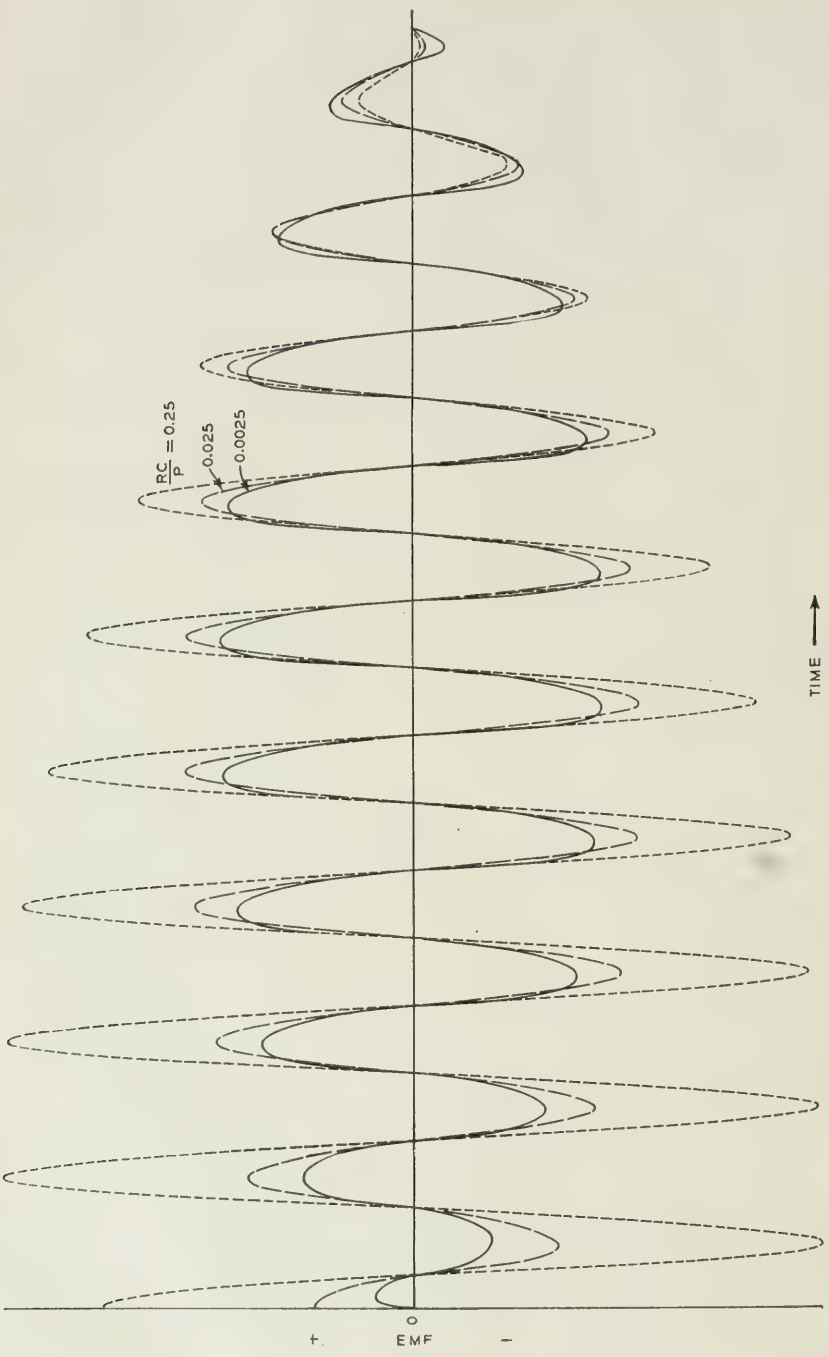


Fig. 8—The wave of Fig. 3*a* is shown here as compressed (1 : 2) by gain changes determined by the impressions of Fig. 6. The smallest time constant gives a result approaching instantaneous compression as in Fig. 5*b*. The next size approaches envelope compression as in Fig. 5*c*. The largest size shown results in considerable distortion, particularly at the beginning. The distortion of the fourth is too large to be shown on the same scale.

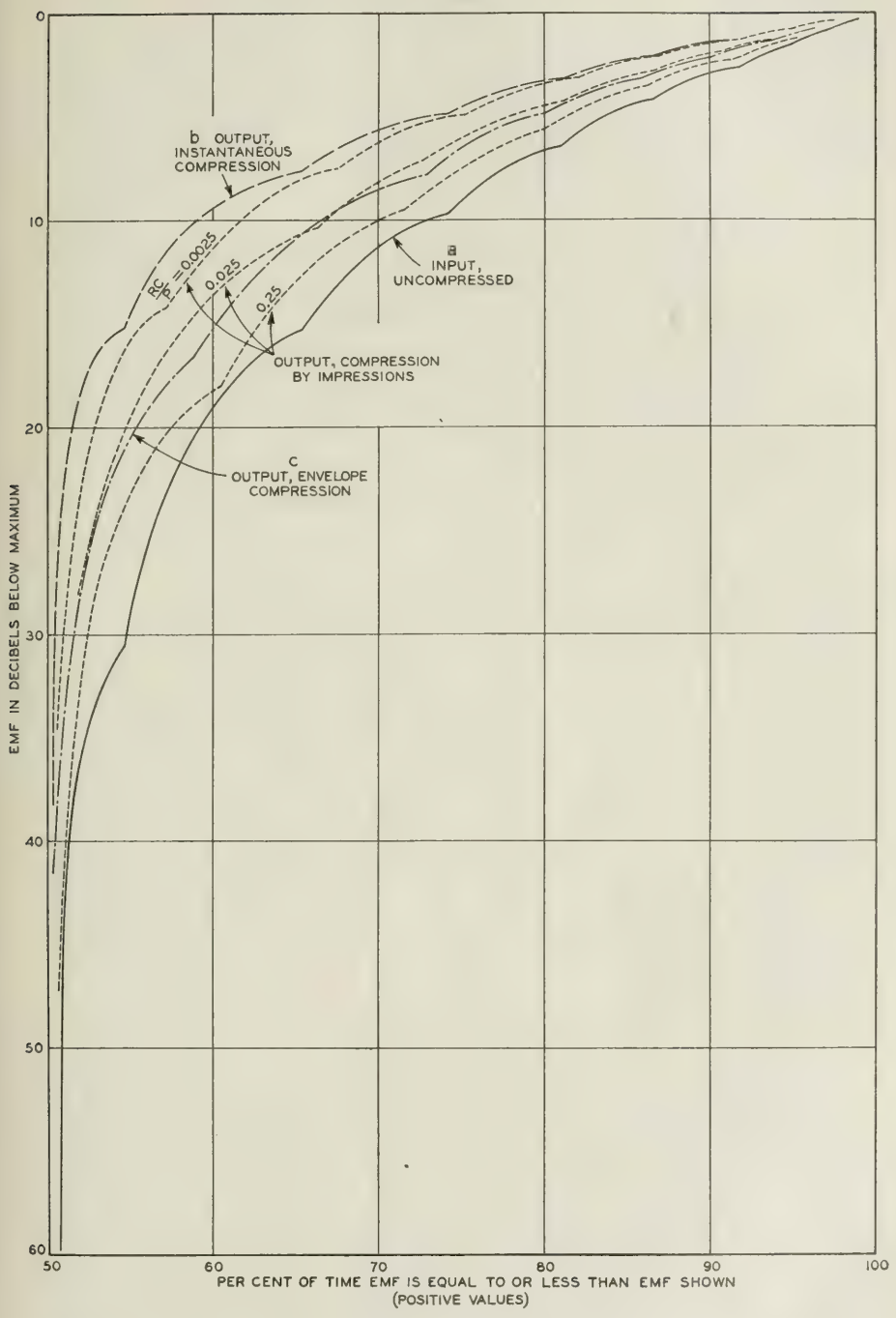


Fig. 9—The amplitude ranges of the compressed waves of Fig. 8 are shown, together with those of Fig. 5. The amount of signal modification (and noise improvement) for any amplitude below maximum may be correlated readily with the time constant.

control ratio. In the case of envelope compression (*c*), the lag causes a reduction in the amount of compression of the instantaneous voltages and the result is seen to be about half way between curves *a* and *b* of Fig. 9. The remaining curves of Fig. 9 are representative of the device [3] used on the long-wave transatlantic radiotelephone circuit.

Volume Control

To avoid both a large range and also the necessity for a continuous record, practical speech amplitude measuring instruments are not directly concerned with either instantaneous, envelope, or impression voltages. Instead a value is determined, corresponding to an impression which is exceeded only a small percentage of the time. This is the principle underlying speech measurements with "volume indicator" type of instruments. In the case of speech, which is much more complex than the simple wave we have discussed, curves like Fig. 9 are steeper, i.e., there are relatively more peaks and a larger range to complicate the problem.

A particular device capable of compressing according to the requirement that the dynamic "volume" range should be reduced, is attained by a combination of several separate range controllers. One is provided to reduce the gain very rapidly when the output volume is too high. A second increases the gain, at a much slower rate, when the impressions formed on the condenser are consistently too low. A third disconnects the condenser from the input when the applied voltage is very small, so that the distortions inherent in change of gain by weak impressions will occur only at times of large and sudden decreases of volume. In the device [4] employed to control volumes applied to a radio transmitter at Norfolk, Virginia, a fourth control provides for rapid partial compression of high peaks, thus improving the modulation. It is unnecessary to re-expand for the purpose of restoring the intelligibility, since the distortion is virtually limited to a change in loudness.

Part 3—Range

Range controllers, like repeaters and attenuators, are limited as to the input range they can accept and the output range they can provide. These limits may be due to thermal noise at the low end and output carrying capacity at the high end. Heretofore, in this paper, the terms "input" and "output" have been purposely left somewhat vague so as to be as general as possible. However, the limits of input and output of a range controller take on particular significance when it is considered that the signal input range may differ both from the

input range of the device, and also from the range over which control is exercised.

This control range may be defined as the difference between the maximum and minimum values of an applied wave over which a device is designed to function in a specific non-linear manner. It is usually expressed in db, and may apply to any measure of the applied signal, such as instantaneous voltage, rms steady-state sine waves, or a dynamic measure such as volume. The values dividing the controlled range from the uncontrolled ranges may be referred to as the "control points."

Certain advantages in some cases have been found from restricting the control range. This is accomplished by placing one or both of the control points inside the useful amplitude range. The position of the control point may be moved arbitrarily over a wide range by putting an ordinary repeater (or attenuator) in tandem with the range controller. A given amount of compression at the high amplitude end of the range gives a real signal-to-noise advantage for a much greater proportion of applied e.m.f.'s than the same amount of compression at the low end of the range. In either case the distortion would be less than that of a full range compressor. When expandors with limited range are used, they are subject to the limitation that variations in the medium are increased, but to a lesser extent than full range expandors.

Part 4—Classification of Range Controllers—Secondary Characteristics

Table I, page 536, suggests how the conceptions of control ratio, time actions and range already discussed might be employed to distinguish a variety of devices. In cases where more than one device is covered by a given control ratio and time action the distinction is that the ranges are different. The names of devices used in this table are those which have been used in the past to distinguish the devices one from another.

Nomenclature

Using the above conceptions of the three primary characteristics, it has been found possible to devise a notation to distinguish all the known devices in this field. As an example of how this proposed system of nomenclature would be applied, Table II, page 537, gives three columns. Column 1 sets forth the arbitrary names that have been used in the past to distinguish certain devices which have come into use. Column 2 gives descriptive names which specify the three fundamental characteristics. In column 3 each device is named by three symbols defining the three fundamental characteristics, and a

TABLE I

CLASSIFICATION OF RANGE CONTROLLERS

Typical Time Actions	Compressors ($r < 1$)		Expandors ($r > 1$)	
	Full Range	Limited Range	Full Range	Limited Range
Instantaneous	Rooter	Peak Chopper, Voltage Limiter	Inverse Rooter, (Squarer)	Voice Operated Relays, Cross-talk Suppressor
Syllabic	Compressor	Limited Range Compressor, Peak Limiter	Expandor	Noise Reducer
Volume	Vogad, Range Reducer	Volume Limiter, Half Vogad	Range Restorer	

classification which tells what the device is designed to do. In this system the numbers preceding the letters specify the input control range in decibels, and the position of a horizontal bar indicates the position of the main signal range with respect to the control range.

The letters specify the time actions and in the case of vogads, where several time actions may be combined, an arbitrary combination of letters would be used. The final numbers specify the control ratio, and in the case of vogads, where this might be different depending on whether the input was increasing or decreasing, both values are given, the former first.

In this system, definitions of time actions are prerequisite and by way of illustration, the following symbols have been used:

I represents instantaneous, meaning very fast adjustment of device
S represents syllabic, meaning moderate speed adjustment of device
V represents volume, meaning a combination of controllers which produces adjustment of device in response to dynamic speech so that the output volume is approximately determined by the input volume.

Secondary Characteristics

In addition to their three primary characteristics, range controllers may have a number of secondary features which are sometimes important. The outstanding ones are:

1. Bias

A *neutral* range controller is one which holds its setting during the quiet periods between words and sentences and which changes its gain

TABLE II

COMPARISON OF NOMENCLATURE FOR RANGE CONTROLLERS

Col. (1) Arbitrary [1]	Col. (2) Systematic	Col. (3) Symbolic
1. Vogad	Full Range 45 db Volume Compressor	45 *VSI* 23–18 Compressor
2. Vogad Combined with Syllabic Compressor	Full Range 45 db Volume Compressor	45 *VSSI* 23–18 Compressor
3. Volume Limiter	High Range 15 db 1 : 5 Volume Compressor	$15\underline{V5}$ Compressor
4. Compandor	Full Range 60 db 2 : 1 Syllabic Compandor	60*S*2 Compandor
5. Noise Reducer	Low Range 20 db 2 : 1 Syllabic Expandor	$\overline{20}S2$ Expandor
6. Limited Range Compressor	High Range 10 db 1 : 2 Instantaneous Compressor	$10\underline{I2}$ Compressor
7. Peak Limiter	High Range 12 db 1 : 5 Syllabic Compressor	$12\underline{S5}$ Compressor
8. Peak Chopper	High Range 6 db 1 : 100 Instantaneous Compressor	$6\underline{I100}$ Compressor
9. Crosstalk Suppressor	Low Range 10 db 10 : 1 Instantaneous Expandor	$\overline{10}I10$ Expandor
10. Rooter and Inverse Rooter	Full Range 70 db 2 : 1 Instantaneous Compandor	70*I*2 Compandor
11. Vodas (Singing Suppressor Relay)		$0I\infty$ Expandor
12. Syllabic Vodas		$0S\infty$ Expandor

only when the waves acting on it differ from those just received. This condition sets a new requirement on the range controller which can usually be met by a combination of control circuits.

A *biased* controller is one which returns to a setting corresponding to some fixed or biased intensity when speech is not passing and adjusts itself each time speech begins. A simple compressor is biased since with no input it generally takes a setting of maximum gain so as to be right for the weakest waves that might be applied in its working range or below the working range. It is also possible to bias a range controller so as to have minimum gain, or any other intermediate value when no waves are applied. An important secondary characteristic is the rate at which the device returns to the desired "bias" point.

Any of the devices listed in the tables might be neutral or biased in either direction, thus increasing the number of possible arrangements.

2. Behavior Outside of Range

For inputs outside the working range of a range controller it is important to provide that the amplification of these waves does not cause them to be modified so as to be out of proportion to output signals in the main range. In some cases this is met by choosing a device which follows the same law all the way to zero current. In others, the device may act as a linear transducer, i.e., with range

factor of one outside the working range. Various other combinations of control ratios can, of course, be employed.

3. Pilot Channel

In all complete compressors some form of pilot channel is necessary to control the re-expansion if this is required. If the gain changes are slow, the pilot channel may include an operator who changes the gain of the receiving device in a manner complementary to that of the sending device based on aural or visual signals. If the gain changes are too rapid for the operator to follow, the receiving gain may be changed automatically.

The pilot channel itself may be a direct or alternating current of variable amplitude or frequency, or in case of carrier or radio, it may be the carrier frequency. In sound reproduction a pilot channel might be a pilot track on the record.

Summary

In an amplitude range control system, the following characteristics must be specified in addition to the usual repeater characteristics, to determine its design and performance:

1. The steady-state control ratio, which determines how much control is obtained and whether restoration can be made automatically or not.
2. The manner in which the output varies with time, following a given change in input.
3. The range over which it is to function.

In specific cases the following should also be considered:

4. The action of the device for inputs outside the working range.
5. If the device is a complete compressor, the type of pilot channel for restoring.
6. The action of the device when signals are removed.

Acknowledgment

The computations used to obtain Figs. 3 to 9, inclusive, were made by Miss Marian Darville.

References

1. "Devices for Controlling Amplitude Characteristics of Telephonic Signals," A. C. Norwine. Presented at A. I. E. E. Pacific Coast Convention, Aug. 9–12, 1938.
2. "Speech Power and Its Measurement," L. J. Sivian, *B. S. T. J.*, Vol. VIII, pp. 646–661.
3. "The Compandor—An Aid Against Static in Radio Telephony," R. C. Mathes and S. B. Wright, *B. S. T. J.*, Vol. XIII, pp. 315–332.
4. "A Vogad for Radio Telephone Circuits," S. B. Wright, S. Doba, A. C. Dickieson. Presented at I. R. E. Convention, June 1938.

Devices for Controlling Amplitude Characteristics of Telephonic Signals*

By A. C. NORWINE

This paper describes a family of devices which automatically respond to signals and control the circuit amplification in such a way as to improve transmission. Their general characteristics are outlined, their differences explained, and some of their applications are listed.

Introduction

THE transmission of speech energy over electrical circuits is attended by the interesting and sometimes difficult problem of preserving the original signal in spite of limitations in the transmission medium. These limitations include load carrying capacity, interference with other service, noise, change in attenuation with time and many others. Because of special limitations it is sometimes desirable to alter the amplitude characteristics of the speech or other signal energy without, of course, materially lowering its intelligibility. In high quality systems the peak voltage from some speech sounds of a given talker may be over 30 db (some 30 times) higher than from his weakest sounds when there is very little inflection in the speech. Loudness changes for emphasis will increase this range of intensities. Ordinary message systems do not have to contend with quite so wide a range of instantaneous voltages from a single talker, but different talkers under extreme terminal conditions produce about a 45 db range of average voltage, which is additive to that for a single talker. Consequently, a voltage range of about 70 db (over 3000 to 1) must be considered for message circuits.

In order to accommodate such ranges of intensity to certain transmission media such as radio links a new family of automatic devices has been developed. In general all of these contain amplifiers or attenuating networks whose loss or gain is changed according to some function of the applied input and which may have a variety of time sequences in their control circuits. It is hoped that by the classification and description of some of these devices their distinguishing characteristics and fields of usefulness will be made somewhat clearer.

* Presented at the Pacific Coast Convention of A. I. E. E. and I. R. E. in Portland, Oregon, August 9–12, 1938.

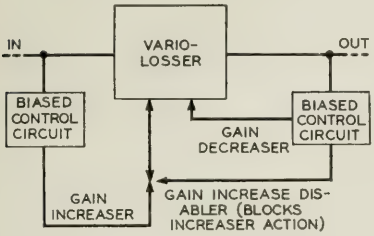
We are to be concerned here principally with those elements allied to the telephonic art, although some applications are to be found in other fields. It is not intended to include those voice operated functions which are essentially switching operations although the distinction in some cases becomes exceedingly fine.

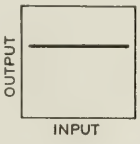
Names of volume controlled devices * which have been used in published papers include vogad,† [2, 3, 4] compandor,‡ [5, 6] and volume limiter.[7] Without direct comparison it may not be obvious how these and similar devices differ. First the apparent similarity of several of these devices will be shown in simple diagrams. Next the more important characteristics of a number of devices will be presented in tabular form, followed by descriptions of the different types. These will then be discussed with particular emphasis on their distinctive qualities, with notes on their variants which have some apparent value.

General Characteristics of Volume Controlled Devices

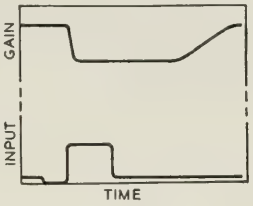
In Figs. 1 to 10 are shown simplified diagrams of some of these devices. While detailed descriptions of them will be deferred till later it may be pointed out that all those shown contain vario-lossers, and all have paths from the main transmission path to control circuits which affect the vario-lossers. A vario-losser usually consists of a balanced pair of vacuum tubes whose gain is changed by varying the grid bias, or of a network of non-linear elements such as copper oxide or silicon carbide whose loss is changed by varying a current through them. In some special cases it may be a mechanically adjusted variable network. The word vario-losser is thus a generic term relating to a circuit whose loss or gain is controllable. A control circuit ordinarily consists of an amplifier and rectifier whose direct current or alternating current output bears a chosen relation to its input. Thus some control circuits are marginal; they produce no control voltage till the input exceeds some critical value, then produce large control voltages for small additional increments of input. These are used, for example, when it is desired to limit the output of a vario-losser to a definite amount. Another type of control circuit produces a current or voltage which is linear with input expressed in decibels. In combination with a vario-losser whose gain is a linear function of control current or voltage one can produce a device whose gain is a linear function in decibels of the input to the control circuit.

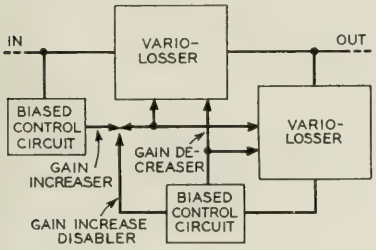
* See the footnote on page 543.
† "*V*olume *O*perated *G*ain *A*djusting *D*evice."
‡ A combination of the names "*Com*pressor" and "Ex*pandor*."

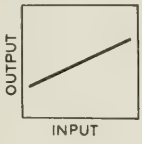




VOGAD
FIGURE 1



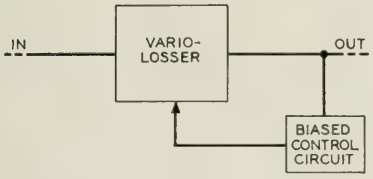


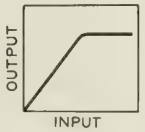


HALF VOGAD
FIGURE 2

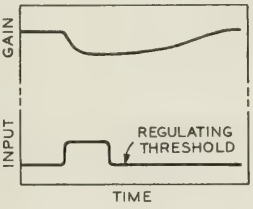
TIME FUNCTIONS SAME AS VOGAD

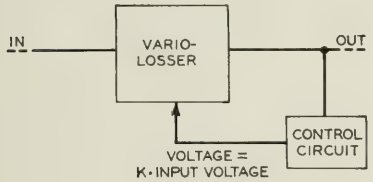
GAIN CHANGES HALF AS GREAT

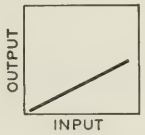




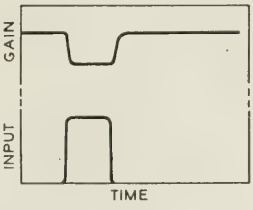
VOLUME LIMITER
FIGURE 3

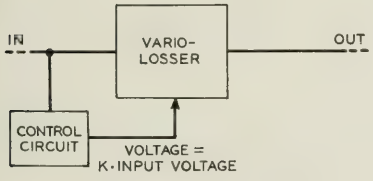


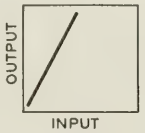




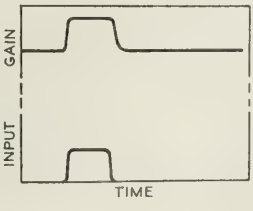
COMPRESSOR
FIGURE 4







EXPANDOR
FIGURE 5

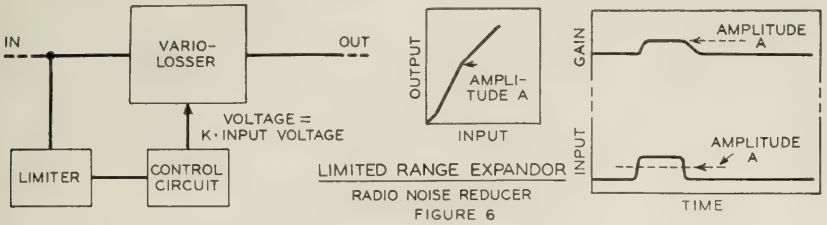


It will be recognized that if the application or removal of the control energy is retarded, the action of the control circuit may be made quite different on transient inputs than on steady state inputs. It will appear later that this is the important distinction between some of the devices to be discussed and that fundamental differences in their functioning are thus brought about.

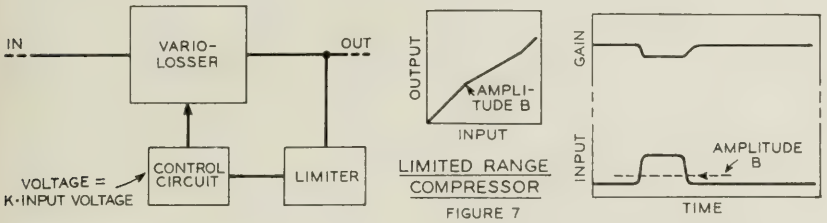
Referring to the figures once more it will be noted that some control devices are connected to the transmission path at the input to the vario-losser. These are known as "forward acting" control circuits. Other controls, connected at the vario-losser outputs, are known as "backward acting" control circuits. This is simply convenient terminology to indicate whether the control energy is progressing in the same direction as the main transmission or is progressing in a backward direction after traversing the main path, usually through a vario-losser. Some backward acting controls function to measure the output of the devices containing them and to make whatever adjustments are required. Others are placed in that position to take advantage of the vario-lossers in the transmission paths, i.e., such controls could be replaced by combinations of forward acting controls and extra vario-lossers.

In Table I, nine of the volume controlled devices * which have been developed for various commercial and experimental uses are listed with the functions of voltage, time, and frequency which are employed to obtain their respective performances. There is, of course, some latitude in the choice of these functions for any one device. Pending more complete description of the different types in the following paragraphs this table should be viewed as illustrating the general character of the different circuits and also the range of the variables which already have been employed. For example, it will be seen that instantaneous voltage of the signal wave, its short time average value, peak power, syllabic variations, and long time average power have all been used as criteria of gain settings in different circuits. Some devices change their adjustments only when critical values or ranges are exceeded, while others vary somewhat with every syllable if speech, for example, is being transmitted. Some are linear transducers to all but low or high amplitudes while others reduce or increase the output range from that at the input. It will be seen that proper choices of times for gain increase and gain decrease in combination

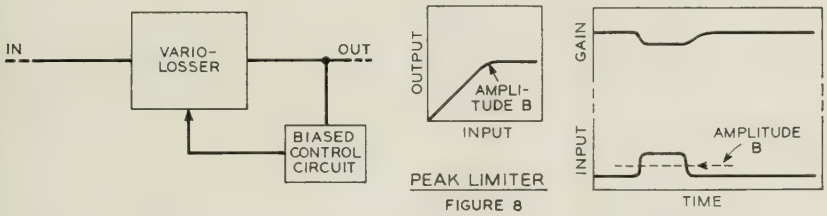
* The names employed do not follow an entirely logical classification, but they are given here because they have had considerable usage. For the same reason the term *volume controlled devices* is used, although to be strictly correct it might better be *sound energy controlled devices*, for example, for not all the devices operate in accordance with *volume* as measured by the well-known class of visual reading meters called *volume indicators*.



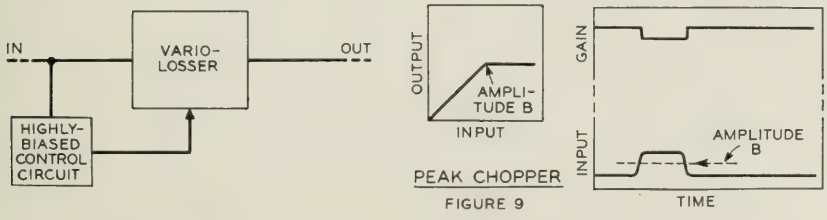
LIMITED RANGE EXPANDOR
RADIO NOISE REDUCER
FIGURE 6



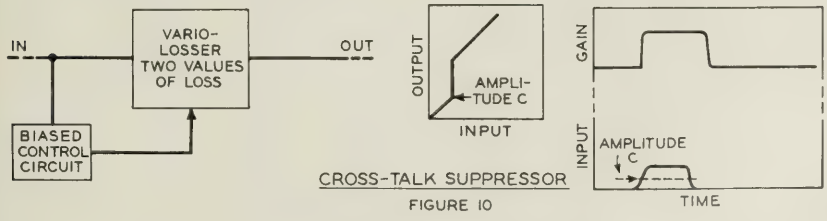
LIMITED RANGE COMPRESSOR
FIGURE 7



PEAK LIMITER
FIGURE 8



PEAK CHOPPER
FIGURE 9



CROSS-TALK SUPPRESSOR
FIGURE 10

TABLE I

CHARACTERISTICS OF VOLUME CONTROLLED DEVICES

Device	Gain Controlled by	Frequency of Adjustment	Time Required for Gain Changes		Ratio of Output Range to Input Range†	Position of Controls	Volume Range Controlled*	Frequency Range Controlled	Frequency Range Causing Control	Part of Input Range Causing Operation
			Gain Increase	Gain Reduction						
1. Vogad	Average volume	Infrequent. Gain fixed between transmissions	After a few words—sometimes 8 to 10 words	After one or more words	Approx. 0	At input and output	Large	Full band transmitted	Full band	All
2. Volume Limiter	Average volume over part of input range	Relatively infrequent. Approaches max. gain in silent periods	Slow and continuous	After one or more words	1 up to operate point. 0 above	At output	Moderate	Full band	Full band	High amplitude only
3. Compandor *a*. Compressor *b*. Expandor	Syllabic variations	Continuous at syllabic rate	On each syllable	On each syllable	1/2 to 1/5 5 to 2	Compressor output expandor, input	Large	Full band	Full band	All
c. Special Compandor	do	do	do	do	do	ditto or expandor control over separate channel	do	High frequency only or multiband	High frequency only or multiband	All or high amplitude only
4. Radio Noise Reducer (Limited Range Expandor)	Syllabic variations over part of input range	Each syllable	On each syllable	On each syllable	2 for low amplitudes. 1 for high	At input	Moderate	Full band	Full band	Variable at low amplitudes only. Fixed gain to high amplitudes

TABLE I (*Continued*)

Device	Gain Controlled by	Frequency of Adjustment	Time Required for Gain Changes		Ratio of Output Range to Input Range†	Position of Controls	Volume Range Controlled*	Frequency Range Controlled	Frequency Range Causing Control	Part of Input Range Causing Operation
			Gain Increase	Gain Reduction						
5. Limited Range Compressor	Syllabic variations over part of input range	Each syllable within the operating range	Fast	Fast	1 up to operate point, then 0 to 1/2, then 1	At input or output	Small	Full band	Full band	High or intermediate amplitudes
6. Peak Limiter	Syllabic peaks	Infrequent	After about one word	On single syllable	1 up to operate point, then approaches 0	At output	Small	Full band	Full band	High amplitude only
7. Peak Chopper	Instantaneous peak voltage	Relatively infrequent	Instantaneous	Instantaneous	1 up to operate point, then 0	At input	Very small	Full band	Full band	High amplitude
8. Crosstalk Suppressor	Voltage exceeding a specified value low in input range	Relatively frequent	Fast	After one or two words	1 except at point of discontinuity	At input	Very small	Full band	Full band	All above specified low value
9. Rooter and Inverse Rooter	Instantaneous voltage	Continuous	Instantaneous	Instantaneous	Ordinarily 1/2 and 2	Integral with vario-losser	Moderate	Full band	Full band	All

* Outside these ranges most of the devices tend to be linear transducers except for higher volumes applied to "Limiters."

† It is important to note that these ranges are measured in the same units as the respective control circuits measure: viz., "volume," "syllabic power," "voltage," etc.

with certain gain control criteria make possible a wide variety of signal altering means to meet different requirements.

Description of Devices in Table I

With this introduction to the combinations of characteristics which are possible it should be less difficult to distinguish between the specific devices discussed in the following paragraphs, which, in addition to describing the devices, contain some comments which should assist in visualizing their forms and their operation.

1. The *vogad* (Fig. 1) is a device which will maintain at its output speech volume [1] which, over a certain range of input, is relatively independent of the speech volume applied to its input and which, in the ideal case, will not change its gain during periods of no speech input. It makes little or no alteration in the ratios of maximum and minimum instantaneous to average voltages of the speech.

2. The *volume limiter* (Fig. 3) is a device which is a linear transducer for all speech volumes up to a critical value, beyond which all input volumes produce essentially the same output volume. It is essentially different from the vogad in that its gain approaches the maximum value when input is removed.

3. The *compandor* (Figs. 4 and 5) is composed of a *compressor* and an *expandor*. A *compressor* is a device whose input-output characteristic on a decibel scale has a slope less than unity * and whose gain or loss is variable under control of the input energy at a time rate which will permit it to follow the syllabic rate of change of speech energy. Similarly, an *expandor* is a device whose input-output curve has a slope greater than unity and whose gain is variable at a syllabic rate under control of the input energy. Thus very shortly after all input is removed the gain of a compressor is maximum and the loss of an expandor is maximum. The reciprocal of the compressor characteristic slope is spoken of as the compression ratio, and the slope of the expandor characteristic is spoken of as the expansion ratio.

4. The *radio noise reducer* [8, 9] (Fig. 6) combines the functions of an expandor which operates in the range of amplitudes where noise and weaker speech sounds lie and a linear transducer which comes into play for all amplitudes exceeding a critical value, which can be set to best suit the atmospheric noise conditions. In other words, the radio noise reducer is a limited range expandor. Inputs which are below the expandor range are subject to transmission at the minimum gain.

5. The *limited range compressor* (Fig. 7) is a device whose operating

* That is, if the input increases by x db the output increases by less than x db.

range includes a region within which compression at a syllabic rate can take place; at other inputs the device is a linear transducer. Its connecting diagram and time functions are the same as those shown in Fig. 5 except that the control circuit contains a limiting device, so that compression takes place in only a portion of its input range, analogous to the action of the limited range expandor of Fig. 6. As a special case the *limited range compressor* may have no linear range above its compression range, thus becoming one type of peak limiter.

6. The *peak limiter* (Fig. 8) is a device whose gain will be quickly reduced and slowly restored when the instantaneous peak power of the input exceeds a predetermined value. The amount of gain reduction is a function of the peak amplitude, and in practice is usually intended to be small to prevent material reduction of the range of intensity of the signal.

7. The *peak chopper* (Fig. 9) is a device which prevents transmission of peak amplitudes exceeding a critical amount, an essential characteristic being that the loss it inserts is completely determined by the instantaneous voltage of the signal. That is, its operating and releasing times are substantially equal to zero.

8. The *crosstalk suppressor* (Fig. 10) is a device which normally presents a prescribed loss to transmission, which loss is removed rapidly when the input amplitude exceeds a certain threshold and is reinserted at a definite time after the input is removed. It reduces low amplitude unwanted currents such as crosstalk but does not affect amplitudes in the useful signal voltage range. This device differs from the limited range expandor in that the time during which the low loss condition is maintained is considerably greater, so the transition from one gain to the other occurs less frequently.

9. A *rooter* is an instantaneous compressor. Such a circuit can be made to produce an output whose instantaneous voltage is, for example, the square root or some similar function of the instantaneous voltage applied to the input. An *inverse rooter* is an instantaneous expandor whose characteristic is complementary to that of the rooter. A combination of *rooter* and *inverse rooter* will reduce the load requirements on a transmission system between the two units but requires that it transmit a wider band of frequencies than that for the original signal, and that it be essentially free from phase distortion. This does not seem to be an attractive arrangement from a commercial viewpoint and is included here simply as an illustration of one of the possible modifications of signal energy. It is not shown in the group of diagrams.

Variants to the Devices Described

In addition to these there are various devices which are essentially modifications of those described. For example, a half-vogad, Fig. 2, may have the same time functions as a vogad, Fig. 1, but the gain changes in the transmission circuit are half as great for the same range of input volumes. Thus in a vogad the range of gain changes in the transmission circuit is equal to the range of input volumes, so that the output volume is the same for all input volumes. In the case of the half vogad the range of gain changes in the transmission path is one-half the range of input volumes, so the output volume range is one-half that of the input. It is also possible to construct a vogad whose output volume range is any desired fraction of the input range. As another example of modification of the devices described, for special applications it may be desirable to incorporate a certain amount of syllabic compression in a vogad.

Communication circuits which have separate paths for oppositely directed transmission between the two terminals are usually operated at such an overall loss that with ordinary terminations there will be little tendency for circulating currents to build up to a "singing" condition. Sometimes there may not be a great deal of margin, however, so that volume controlled devices added to such circuits must add loss at some point to counterbalance whatever gain is put in at some other point. Thus a vogad inserted at the transmitting side of one terminal of such a circuit to amplify speech energy from weak talkers must be supplemented by a "reverse vogad" in the receiving side of the circuit. The reverse vogad is simply another vario-losser which is operated upon by the vogad control circuit in such a way that it always has a loss numerically equal to the gain of the vogad. Any vogad gain will be compensated by the reverse vogad loss, so no greater tendency to sing will be effected by the addition of the combination to the circuit. In like manner half vogads must be used with compensating reverse half vogads.

Combinations of some of the devices also have interesting characteristics. For example, a combined radio noise reducer and peak limiter at the receiving end of a circuit would suppress noise and would also reduce the amplitude of excessively high amplitude signals. Likewise, a vogad, compressor, and peak chopper in tandem in the order named could be made to reduce the range of input signals by a very large amount for transmission over a medium having only a small range between noise and maximum permissible signal. In this case it would be practically impossible to recover the original signal range at

the receiving terminal of the medium, but the intelligibility of speech over such a system has been shown in the laboratory to be good.

Special compandors for high quality service may require compression and expansion which vary with frequency. The exact characteristics will depend upon band width, program material and transmission medium. For transmission media in which the noise reproduced at the receiving end is principally at the higher frequencies an unusual effect is obtained if the usual variety of compandor is used. Low frequencies unaccompanied by high frequencies will cause a gain change in compressor and expandor, thus changing the background of high-frequency noise which is not masked by the low-frequency signal energy. The resulting swishing noise has been given the onomatopœic name of "hush-hush effect." To avoid this, recourse may be had to split band compandors in which the compression and expansion is done only at high frequencies or separately for low and high frequencies. The successful application of the latter method is, however, more difficult than it appears from its simple description.

Distinguishing Characteristics

It is important to distinguish between the half vogad, Fig. 2, and the compressor, Fig. 4. As shown in Table I the latter operates on syllabic variations and the former on the average volume of the input. Thus the half vogad reduces the range of output volumes to one-half that at the input while the compressor reduces the range of syllabic power at its output to one-half that at the input. In other words, the compressor reduces the ratio of peak to average power on constant volume speech, while the half-vogad simply adjusts for that volume and does not alter the peak ratio. There is, of course, the additional important difference that the half-vogad retains its gain setting during silent periods while the compressor, by virtue of having followed the syllabic power, has its maximum gain during silent periods.

Volume limiters, Fig. 3, may be mistaken for vogads, Fig. 1, because during speech input above a certain value the two may produce the same output volume. They both employ something like a measurement of average power over periods longer than a syllable to determine their gain settings. The important difference is that a vogad retains its gain setting when speech currents are not present, while a volume limiter approaches its maximum gain during such periods. In terms of the output resulting from a range of input volumes there is another important difference if the volume limiter operates over only part of the input range: the vogad reduces the width of the distribution curve of volumes to a very small value, while the volume limiter moves all the

area under the distribution curve above a certain point to the region near that point, which is its limiting volume. This is illustrated in Fig. 11, in which the calculated modifications of a volume distribution by a vogad and by a volume limiter are shown. In the cases "without volume control" and "with a vogad" the distributions are normal, and the standard deviation, σ, has its usual statistical significance. With a volume limiter, only volumes above the limiting volume are affected,

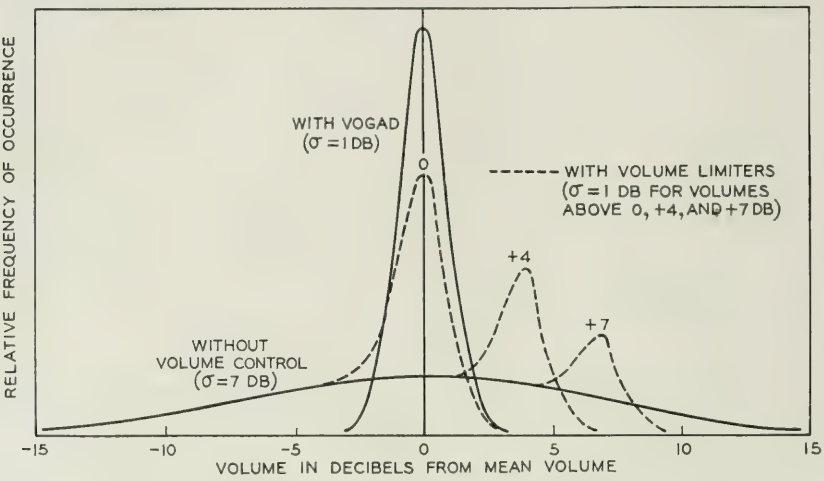


Fig. 11—Modification of volume distribution by use of a vogad or a volume limiter.

and these higher volumes are redistributed according to a normal law whose standard deviation is 1 decibel, as stated in the figure.

It is also important to distinguish between a peak limiter and a peak chopper, Figs. 8 and 9. Naturally they resemble one another since they are intended to permit transmission of signals at higher average amplitudes without excessive loading of transmission circuits. However, they are intended for different classes of service and hence are not interchangeable except in some borderline cases. For the highest grade of transmission harmonic production must be negligible and the reduction in amplitude range of signals small and infrequent. Gain changes must be smooth, though rapid enough to compensate for practically any input wave to be expected. These characteristics are found in the peak limiter now being furnished for use on program networks and radio transmitters.[10, 11] For services in which it is desirable to maintain the signal energy at a high value to over-ride noise and in which harmonic distortion must be kept low a peak limiter with somewhat smaller time constants may be used. A high ratio limited range compressor might be suitable in this instance. This device would lower its gain a little more quickly on excessive

inputs, and it would also reinsert its gain much more quickly; it would affect the naturalness of the sound of the signal more than the slower peak limiter but it would also cause the signal to over-ride noise somewhat better. In a third variety of service the harmonic distortion introduced by a limiter is a secondary matter, the prime consideration being that the peak amplitude of the signal shall not exceed a specified value. This may be because higher amplitude signals would produce a tremendous increase in distortion or crosstalk into other channels or would damage expensive equipment farther along in the circuit. For these cases we may use the fastest possible type of limiter, the peak chopper, which simply cuts off any peak exceeding a certain value.

The crosstalk suppressor, Fig. 10, is a splendid example of the fine distinction between volume controlled and voice operated switching devices. This device has been described, but in the present state of the art its time functions have not been definitely fixed. If the characteristic of loss versus input is made steep enough and the speed of operation fast enough it will sound like a switching circuit and may in fact be replaced by a relay-switched attenuating network. If made somewhat slower and given a smaller slope of loss versus input it approaches the limited range expandor or noise reducer.

Applications and Expected Advantages

It may be of interest to give some approximate figures on the magnitudes of the advantages to be obtained by the use of some of these devices. It will be understood that the values to be given are simply illustrative, some having been obtained from field service on particular models and some from tests on laboratory equipment under special conditions.

Vogads appear to be most useful in such circuits as transoceanic radio connections, where it is important to properly operate the terminal switching equipment and to transmit over the radio circuit speech energy from loud and weak talkers equally well. It is essential in such cases that noise should not be increased in amplitude during speech pauses, hence the gain retaining feature of the vogad. On such a circuit a vogad will reduce a 45 db volume range to about 2 to 4 db. This is equivalent to expert manual volume control.

Volume limiters are in use at the present time to prevent peaks of speech energy in carrier circuits from "splashing" into telegraph channels.[7] Some 5 to 10 db limiting is allowed on loudest talkers, which causes little degradation of the speech channels but makes possible the use of telegraph on the same carrier system. There is no

wide-spread use of volume limiters in point-to-point radio service so far, but in cases in which there is no disadvantage in raising noise in silent periods in speech, such as in push-to-talk installations, proper transmitter loading can be obtained with volume limiters fairly cheaply.

One commercial model peak limiter, used as part of a program amplifier [10, 11] is capable of introducing a considerable amount of compression without overloading on peaks, but for the preservation of adequate program volume range it is being recommended that only 3 db peak limiting be allowed. This, of course, reduces the range of intensity of the program, but from the standpoint of the listeners it is equivalent to doubling the transmitted power or obtaining the same signal-to-noise ratio with half the transmitted power.

Limited range compressors might be used either on land lines to insure full loading or on radio links whose fading is too severe to permit the use of normal compandors. There is no commercial application of either sort at the present time. Peak choppers are, however, used on some high power radio transmitters which might otherwise be temporarily disabled by high peaks in the signal being transmitted.

The chief usefulness of compandors is on radio links in which the transmission of a compressed signal with subsequent expansion permits operation through higher noise or with lower transmitter power. On a long-wave transatlantic radio telephone circuit a compandor with 40 db range has been shown to allow an increase in noise of some 5 db before reaching the commercial limit.[5] With smaller amounts of noise the noise advantage of the compandor approaches half its range in decibels. This benefit is sometimes applied to a reduction of transmitter power.

Radio noise reducers have been used to advantage in connection with short-wave ship-to-shore and transoceanic radio telephone service. In the former, routine transmission rating is given on a judgment basis using a merit scale from 1 to 5, 5 being practically perfect transmission and 1 so poor that intelligibility is very close to zero. It will then be seen that the observed improvement of ½ to 1 point in transmission rating due to the noise reducer is of considerable importance. Perhaps more graphic figures are those for transoceanic service, where the reduction of noise in the receiving path not only reduces the noise heard by the listener but also improves the voice operated switching with the indirect result that at times receiving volume increases of 5 to 15 db are realized.[9]

As has been noted, the radio noise reducer is a special use of an expandor alone. There are also two interesting applications for a

compressor alone. The first, which uses a fairly high ratio of compression, has been mentioned as one type of peak limiting device. The second, using a moderate ratio of compression, is in connection with announcing systems for use in very noisy locations. Its effect is to amplify weak sounds more than strong sounds, which considerably improves the intelligibility through high noise. For quiet locations it is of less value, since the speech sounds lose some of their naturalness in this process.

Conclusion

In the course of developing various types of the volume controlled devices which have been described means have been worked out for providing almost any combination of time constants, range of control, and other characteristics which may be required. Some devices for which there were specific commercial applications or useful functional characteristics for experimental work have been constructed, with resulting advantages which have been briefly mentioned. There remain many possible ways to alter the characteristics of signal energy such as speech to which these methods are applicable and which await the special needs of new transmission problems.

Bibliography

1. C. C. I. F. White Book, 1 bis, pp. 77, 343.
2. C. C. I. F. White Book, 1 bis, pp. 251–3.
3. "A Vogad for Radio Telephone Control Terminals," S. Doba, Jr., *Bell Laboratories Record*, Oct. 1938, Vol. 17, No. 2, pp. 49–52.
4. "A Vogad for Radio Telephone Circuits," S. B. Wright, S. Doba, Jr., and A. C. Dickieson, Presented at *I. R. E.* Convention in New York, June 18, 1938; to be published in *Proc. I. R. E.*
5. "The 'Compandor'—An Aid Against Static in Radio Telephony," R. C. Mathes and S. B. Wright, *Elec. Engg.*, 1934, Vol. 53, No. 6, pp. 860–6; *Bell Sys. Tech. Jour.*, July 1934, Vol. 13, No. 3, pp. 315–32.
6. "The Voice Operated Compandor," N. C. Norman, *Com. and Br. Engg.*, Nov. 1934, Vol. 1, No. 1, pp. 7–9; *Bell Lab. Record*, Dec. 1934, Vol. 13, No. 4, pp. 98–103.
7. "Volume Limiter Circuits," G. W. Cowley, *Bell Lab. Record*, June 1937, Vol. 15, No. 10, pp. 311–15.
8. "A Noise Reducer for Radio Telephone Circuits," N. C. Norman, *Bell Lab. Record*, May 1937, Vol. 15, No. 9, pp. 702–7.
9. "Radio Telephone Noise Reduction by Voice Control at Receiver," C. C. Taylor, *Elec. Engg.*, Aug. 1937, Vol. 56, No. 8, pp. 971–4, 1011; *Bell Sys. Tech. Jour.*, Oct. 1937, Vol. 16, No. 4, pp. 475–86.
10. "Higher Volumes Without Overloading," S. Doba, Jr., *Bell Lab. Record*, Jan. 1938, Vol. 16, No. 5, pp. 174–8.
11. "A Volume Limiting Amplifier," O. M. Hovgaard, *Bell Lab. Record*, Jan. 1938, Vol. 16, No. 5, pp. 179–84.

For the sake of completeness the following references are included, although no allusion has been made to them under the specific device-names used in this paper.

12. "Über automatische Amplitudenbegrenzer," H. F. Mayer, *E. N. T.*, 1928, Vol. 5, No. 11, pp. 468–72.

13. "High Quality Radio Broadcast Transmission and Reception," Stuart Ballantine, *Proc. I. R. E.*, May 1934, Vol. 22, No. 5, pp. 564–629.
14. "Expanding the Music," A. L. M. Sowerby, *Wireless World*, Aug. 24, 1934, Vol. 35, No. 8, pp. 150–2.
15. "Extending Volume Range," *Radio Engg.*, Nov. 1934, Vol. 14, No. 11, pp. 7–9, 13.
16. "Amplitudenabhängige Verstärker," W. Nestel, *E. T. Z.*, 1934, Vol. 55, No. 36, pp. 882–4.
17. "An Automatic Volume Expandor," W. N. Weeden, *Electronics*, June 1935, Vol. 8, No. 6, pp. 184, 5.
18. "Die Arbeitsweise der selbsttätigen Regelapparaturen," H. Bartels and W. G. Ulbricht, *E. N. T.*, 1935, Vol. 12, No. 11, pp. 368–79.
19. "Practical Volume Expansion," C. M. Sinnett, *Electronics*, Nov. 1935, Vol. 8, No. 11, pp. 428–30, 446.
20. "Light-bulb Volume Expandor," *Electronics*, Mar. 1936, Vol. 9, No. 3, p. 9.
21. "Simplified Volume Expansion," W. N. Weeden, *Wireless World*, Apr. 24, 1936, Vol. 38, No. 17, pp. 407–8.
22. "Practical Volume Compression," L. B. Hallman, Jr., *Electronics*, June 1936, Vol. 9, No. 6, pp. 15–17, 42.
23. "Notes on Contrast Expansion," Gerald Sayers, *Wireless World*, Sept. 18, 1936, Vol. 39, No. 12, p. 313.
24. "Contrast Amplification: A New Development," W. N. Weeden, *Wireless World*, Dec. 18, 1936, Vol. 39, No. 25, pp. 636–38.
25. "Overmodulation Control and Volume Compression with Variable-mu Speech Amplifier," W. B. Plummer, *Q. S. T.*, Oct. 1937, Vol. 21, No. 10, pp. 31–33.
26. "Limiting Amplifiers," John P. Taylor, *Communications*, Dec. 1937, Vol. 17, No. 12, pp. 7–10, 39–40.
27. "Low Distortion Volume Expansion Using Negative Feedback," B. J. Stevens, *Wireless Engr.*, Mar. 1938, Vol. 15, No. 174, pp. 143–9.
28. "Distortion Limiter for Radio Receivers," M. L. Levy, *Electronics*, Mar. 1938, Vol. 11, No. 3, p. 26.
29. "Automatic Modulation Control," L. C. Waller, *Radio*, Mar. 1938, No. 227, pp. 21–6, 72, 74.
30. "An AVE Noise Silencer Unit," McMurdo Silver, *Radio News*, May 1938, Vol. 20, No. 11, pp. 46, 55.

The Exponential Transmission Line *

By CHAS. R. BURROWS

The theory of the exponential transmission line is developed. It is found to be a high pass, impedance transforming filter. The cutoff frequency depends upon the rate of taper.

The deviation of the exponential line from an ideal impedance transformer may be decreased by an order of magnitude by shunting the low impedance end with an inductance and inserting a capacitance in series with the high impedance end. The magnitudes of these reactances are equal to the impedance level at their respective ends of the line at the cutoff frequency.

For a two-to-one impedance transformer the line is 0.0551 wavelengths long at the cutoff frequency. For a four-to-one impedance transformer the line is 0.1102 wave-lengths long at the cutoff frequency, etc.

The results have been verified experimentally. Practical lines 50 meters and 15 meters long have been constructed which transform from 600 to 300 ohms over the frequency range from 4 to 30 mc. with deviations from the ideal that are small compared with the deviations from the ideal of commercial transmission lines, either two-wire or concentric.

When an exponential line is used as a dissipative load of known impedance instead of a uniform line it is possible to approach more nearly the ideal of constant heat dissipation per unit length. This makes it possible to use a shorter line.

THE exponential line may be defined as an ordinary transmission line in which the spacing between the conductors (or conductor size) is not constant but varies in such a way that the distributed inductance and capacitance vary exponentially with the distance along the line. That is, the impedance ratio for two points a fixed distance apart is independent of the position of these two points along the line. A disturbance is propagated down an exponential transmission line in the same manner as it would be down a uniform line with the additional effect that the voltage is increased by the square root of the change in impedance level and the current is decreased by the reciprocal of this quantity.

The exponential line has the properties of a high pass impedance transforming filter. The cutoff frequency depends upon the rate of

* Presented before joint meeting of U. R. S. I., and I. R. E., Washington, D. C., April 1938. Published in *Communications*, October 1938.

taper. As the frequency is increased the transfer constant * approaches the propagation constant of the equivalent uniform line. At sufficiently low frequencies the only effect of the line is to connect the input to the load.

Above cutoff the magnitudes of the characteristic impedances at any point are approximately equal to the nominal characteristic impedance * at that point but their phase angles (in radians) differ by an amount which at the higher frequencies is equal to the cutoff frequency divided by the frequency in question. The ratio of input impedance to the input impedance level * of an exponential line terminated in a resistance equal to the impedance level at the output always remains within the range from $1 - f_1/f$ to $1/(1 - f_1/f)$ for frequencies, f, greater than the cutoff frequency, f_1. For a 2 : 1 transformation this means that the input impedance remains within $\pm$ 6 per cent of the desired value for all frequencies above that for which the line is a wave-length long. For a 4 : 1 transformation under the same conditions the irregularities are twice as great.

A transforming network having deviations from the ideal of the order of $\pm (f_1/f)^2$ may be made by connecting an inductance in parallel with the low impedance terminal and a capacitance in series with the high impedance terminal. The magnitudes of these reactances are such that their impedances are equal to the impedance levels of the line at their respective ends at the cutoff frequency. Or expressed in another way the capacitance is equal to $2/(k - 1)$ times the electrostatic capacitance of the line and the inductance is the same factor times the total loop inductance of the line where k is the impedance transformation ratio of the line.

Figure 1 shows the theoretical input impedance-frequency characteristics for 2 to 1 step-up and step-down exponential lines. Curve 1 is for the line with a resistance termination. At low frequencies the input impedance is equal to the load impedance while at high frequencies the line approaches an ideal transformer. Curve 2 is the input impedance of the line terminated with the appropriate resistance-reactance combination. The improvement in the input impedance characteristic for frequencies above the cutoff frequency is evident. At the lower frequencies the input impedance does not approach the terminal reactance but approaches the reactance of the capacitance of the line in parallel with the series terminal capacitance for the step-up line and the reactance of the inductance of the line in series with the shunt terminal inductance for the step-down line. The improvement is not as great as apparent from the figures because the phase angle is

* See appendix for definition of terms.

not improved proportionally. This is easily remedied by completing the impedance transforming network with the appropriate reactance at the input. The resulting input impedance is shown in curve 3. In the "pass" frequency range the maximum reactive component is of

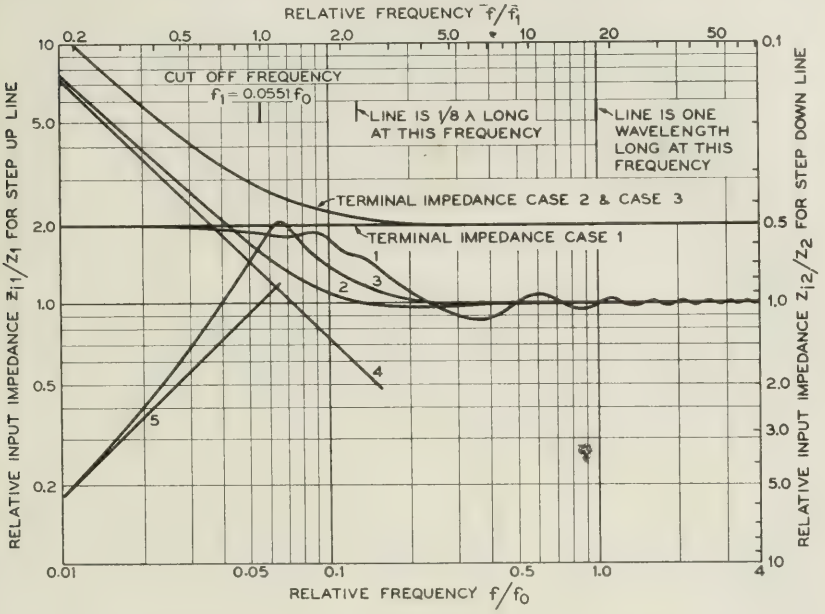


Fig. 1—Input impedance characteristics of 1 : 2 exponential lines. Left ordinate scale refers to step-up line. Right ordinate scale refers to step-down line.

Curve 1—Resistance termination.

Curve 2—With capacity equal to twice the electrostatic capacity of the line in series with the same resistance, $Z_2 = Z_2(1 - jf_1/f)$, for step-up line, or with an inductance equal to twice the total inductance of the line in shunt with the same resistance, $Z_1 = Z_1/(1 - jf_1/f)$, for step-down line.

Curve 3—Termination as for curve 2 with inductance equal to twice the total inductance of the line in parallel with input to the line, $Z_{i1} = Z_1/(1 - jf_1/f)$, for step-up line, or termination as for curve 2 with capacity equal to twice the static capacity of line in series with input to the line, $Z_{i2} = Z_2(1 - jf_1/f)$.

Curve 4—Asymptotic value of impedance of capacity of line in parallel with termination for case 2 for step-up line, or asymptotic value of impedance of inductance in series with termination for case 2 for step-down line.

Curve 5—Impedance of shunt inductance added at input for case 3 for step-up line, or impedance of capacity added in series at input for case 3 for step-down line.

the same order of magnitude as the deviation of the impedance from the ideal.

Besides its application as an impedance transforming network, the exponential line may be used as a "resistance" load of constant known impedance that has a high capability for dissipating power. As such it is capable of dissipating more power in the same length of line than

the uniform line. If x is the maximum attenuation in nepers that can be obtained with a uniform line without overheating, the same length of exponential line will have an attenuation of $(e^{2x} - 1)/2$ nepers.

Exponential lines of the proper length have properties similar to half-wave and quarter-wave uniform lines. The input impedance of an exponential line an even number of quarter wave-lengths long is equal to the load impedance times the impedance transformation ratio of the line. When the length of the line differs from an odd multiple of a quarter wave-length by an amount that depends upon the frequency and load impedance, the input impedance is equal to the product of the terminal impedance levels divided by the load impedance.

Mathematical Formulation

The telegraph equations for the exponential line may be solved by the methods employed in the problem of a uniform line. The resulting equations for the voltage and current at any point along the line are

$$v_x = Ae^{-\left(\Gamma-\frac{\delta}{2}\right)x} + Be^{+\left(\Gamma+\frac{\delta}{2}\right)x} = Ae^{-\left(\Gamma-\frac{\delta}{2}\right)x}\left[1 + \frac{B}{A}e^{2\Gamma x}\right] \quad (1)$$

and

$$i_x = \frac{A}{Z_0}\frac{\Gamma - \frac{\delta}{2}}{\gamma}e^{-\left(\Gamma+\frac{\delta}{2}\right)x} - \frac{B}{Z_0}\frac{\Gamma + \frac{\delta}{2}}{\gamma}e^{+\left(\Gamma-\frac{\delta}{2}\right)x}$$

$$= \frac{A}{Z_0}\frac{\Gamma - \frac{\delta}{2}}{\gamma}e^{-\left(\Gamma+\frac{\delta}{2}\right)x}\left[1 - \frac{B}{A}\frac{\Gamma + \frac{\delta}{2}}{\Gamma - \frac{\delta}{2}}e^{2\Gamma x}\right], \quad (2)$$

where

$\delta = \dfrac{\log_e z/z_0}{x} = \dfrac{\log_e y_0/y}{x} = \dfrac{\log_e Z/Z_0}{x}$ is the rate of taper,

$Z_x = \sqrt{z/y} = Z_0 e^{\delta x}$ is the *surge* or *nominal characteristic impedance* of the exponential line at the point x which is equal to the characteristic impedance of the uniform line that has the same distributed constants as this line has at the point x,

$\gamma = \sqrt{zy} = \sqrt{z_0 y_0}$ is the *propagation constant* of any uniform line that has the same distributed constants as this line at any point. It is independent of the point along the line to which it is referred, and

$\Gamma = \sqrt{\gamma^2 + \delta^2/4} = \alpha + j\beta$ is the *transfer constant* of the exponential line.

$+\gamma$ and $+\Gamma$ refer to the values of the indicated roots that are in the first quadrant.

If these equations are compared with those for a uniform transmission line it is found that the *propagation constant* is $\Gamma - \delta/2$ for voltage waves traveling in the positive x direction and $\Gamma + \delta/2$ for voltage waves traveling in the negative x direction. For current waves the corresponding *propagation constants* are $\Gamma + \delta/2$ and $\Gamma - \delta/2$. In the terminology of wave filters, Γ is the *transfer constant* and δ is the impedance *transformation constant.* $\delta/2$ is the *voltage transformation constant* and $-\delta/2$ is the *current transformation constant.* The real and imaginary parts of Γ, α and β are the *attenuation* and *phase constants* respectively.

An important parameter is

$$\nu = j\frac{\delta}{2\gamma},$$

which for a non-dissipative line is the ratio of the cutoff frequency to the frequency, as can be seen if we write the transfer constant as

$$\Gamma = \gamma\sqrt{1 - \nu^2},$$

where the indicated root is in the fourth quadrant. For a non-dissipative line ν is real and the transfer constant is real or imaginary depending on whether ν^2 is greater than or less than unity. Hence the exponential line is a high pass filter whose cutoff frequency, f_1, is that frequency for which $\nu = \pm 1$. The transfer constant is then less than that for a uniform line by the factor $\sqrt{1 - \nu^2}$ so that both phase velocity and wave-length are larger for the exponential line than for the uniform line by the reciprocal of this factor.

If we terminate this line at $x = l$ with an impedance $Z_l = v_l/i_l$, the ratio of the reflected to direct voltage wave is found to be

$$\frac{B}{A} = -\frac{1 - (Z_l/Z_l)(\sqrt{1 - \nu^2} + j\nu)}{1 + (Z_l/Z_l)(\sqrt{1 - \nu^2} - j\nu)}\, e^{-2\Gamma l}, \tag{3}$$

where the coefficient of the exponential is the *voltage reflection coefficient.*

There will be no reflection if

$$Z_l = Z_l/(\sqrt{1 - \nu^2} + j\nu) = Z_l^{+}, \tag{4}$$

which becomes $Z_l e^{-i\sin^{-1}\nu}$ above the cutoff frequency for non-dissipative lines. This is the magnitude of the forward-looking *characteristic impedance* at $x = l$ as can be seen by dividing the first term of (1) by the first term of (2). Curve 1 of Fig. 2 gives the charac-

teristic impedance of a non-dissipative exponential line looking toward the high impedance end as a function of frequency. At infinite frequency the characteristic impedance is a resistance equal to the nominal characteristic impedance but as the frequency is decreased the phase angle of the characteristic impedance changes so that its locus

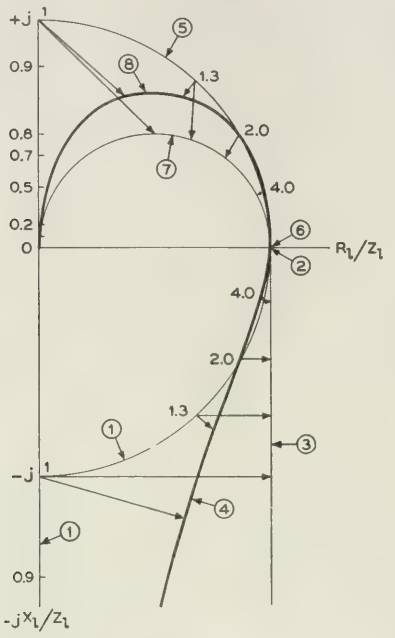


Fig. 2—Impedance diagram comparing the forward looking characteristic impedance with various terminal impedances. The numbers give the frequency relative to cutoff. The arrows are the vectors $Z_l - Z_l^+$ which are a measure of the magnitude of the reflection.

A. Step-up line.

Curve 1—Forward looking characteristic impedance,

$$Z_l^+ = Z_l e^{-i \sin^{-1}(f_1/f)}, \qquad f > f_1,$$

$$Z_l^+ = Z_l[-j(f_1/f)(1 + \sqrt{1 - f^2/f_1^2})], \qquad f_1 > f;$$

Curve 2—Resistance termination, $Z_l = Z_l$;
Curve 3—Capacity resistance termination, $Z_l = Z_l(1 - jf_1/f)$;
Curve 4—Capacity, resistance and inductance termination adjusted for no reflection at twice the cutoff frequency and at infinite frequency;

B. Step-down line.

Curve 5—Forward-looking characteristic impedance,

$$Z_l^+ = Z_l e^{+j \sin^{-1}(f_1/f)}, \qquad f > f_1,$$

$$Z_l^+ = Z_l[+j(f_1/f)(1 - \sqrt{1 - f^2/f_1^2})], \qquad f_1 > f;$$

Curve 6—Resistance termination $Z_l = Z_l$;
Curve 7—Inductance resistance termination $Z_l = Z_l/(1 - jf_1/f)$;
Curve 8—Inductance, resistance and capacity termination adjusted for no reflection at twice the cutoff frequency and at infinite frequency.

is the circular arc. At and below cutoff it is a pure reactance. If the load is a resistance equal to the nominal characteristic impedance at the terminal as indicated at 2 of Fig. 2, there will be no reflection at infinite frequency, but as the frequency is lowered there will be an increasing impedance mismatch with its accompanying reflected wave.

This reflection may be materially reduced by inserting a condenser in series with the resistance load as shown by curve 3. Further improvement results from more complicated networks. Curve 4 shows the effect of adding an inductance in shunt with the resistance load of the resistance-capacitance combination. The arrows indicate the resulting impedance mismatch which is a measure of the reflected wave.

The characteristic impedance looking toward the low impedance end is the inverse of that looking in the other direction as shown by curve 5. Shunting the resistance load with an inductance gives the impedance curve 7. Adding a capacitance element gives curve 8.

Division of (1) by (2) and substitution of the result of (3) gives the following ratio for the impedance looking into the line at the point x to the impedance level at that point,

$$\frac{Z_x}{Z_x} = \frac{K(\sqrt{1-\nu^2} - j\nu) + 1 + [K(\sqrt{1-\nu^2} + j\nu) - 1]e^{-2\Gamma(l-x)}}{K + j\nu + \sqrt{1-\nu^2} - [K - \sqrt{1-\nu^2} + j\nu]e^{-2\Gamma(l-x)}}, \quad (5)$$

where $K = Z_l/Z_l$ is the ratio of the load impedance to the impedance level at the terminal. Here as before the indicated root is in the fourth quadrant.

Network Characteristics

Three parameters are required to specify the characteristics of an exponential line of negligible loss: (1) the cutoff frequency, f_1, (2) the length of the line which is perhaps best specified as the frequency, f_0 = velocity of light/length of line, for which the line is one wavelength long, and (3) the impedance level at some point along the line. We will designate the impedance levels at the low and high impedance ends of the line by Z_1 and Z_2 respectively, and their ratio Z_2/Z_1 by k.

When the line is terminated in a resistance equal to the impedance level at the output (5) reduces to

$$\frac{Z_1}{Z_1} = k^{\cos 2\zeta} e^{2j\zeta} \frac{1 + j \tan \zeta \, k^{-\cos 2\zeta}}{1 - j \tan \zeta \, k^{\cos 2\zeta}}, \qquad \nu > 1, \quad (6)$$

$$\frac{Z_1}{Z_1} = e^{-2j\xi} \frac{1 + \tan \xi \, e^{j\left(2\xi + \frac{\pi}{2} - \eta\right)}}{1 + \tan \xi \, e^{j\left(-2\xi - \frac{\pi}{2} - \eta\right)}}, \qquad \nu < 1. \quad (7)$$

for frequencies below and above cutoff respectively. Here $\eta = -j2\Gamma l$ is twice the electrical length of the line in radians, $\sin 2\zeta = 1/\nu$, $\sin 2\xi = \nu$ and $\cos 2\xi$ is ratio of the electrical length of the line to that of a uniform line of the same physical length. For the step-down line the corresponding ratios are the reciprocal of the above expressions. These ratios are plotted in Fig. 1.

When $f \to 0$, $Z_1 = kZ_1 = Z_2 = Z_2$ and the only effect of the line is to connect the load to the input. Above cutoff the magnitude of the input impedance oscillates about the nominal characteristic impedance and the phase angle oscillates about the value $-2\xi(\approx -f_1/f$ for $f \gg f_1)$ which goes from $-\pi/2$ to 0 as the frequency increases indefi-

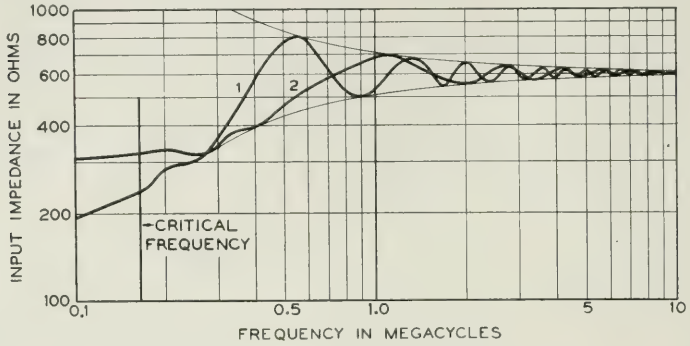


Fig. 3—Input impedance characteristics.
Curve 1—150 : 600 ohm line, 100 meters long.
Curve 2—300 : 600 ohm line, 200 meters long.
Both lines have the same rate of taper.

nitely from cutoff. The variation of the input impedance with frequency is shown for two lines of different length but the same rate of taper in Fig. 3. The magnitude of the oscillations depends only on the rate of taper and decreases with increase in frequency. The impedance varies between $(1 + f_1/f)$ and $1/(1 + f_1/f)$. The positions of the maxima and minima, however, are determined by the length of the line. They occur respectively at those frequencies for which the line is approximately 1/8 of a wave-length more than an even or an odd number of quarter wave-lengths long. The phase angle is usually negative but has a small positive value when the line is approximately a half wave-length long.

The locations of these maxima and minima are the same as would result from terminating a uniform line in an impedance whose magnitude is the same as the characteristic impedance but has a small reactive component. This suggests adding a compensating reactance

to the resistance load. From (3) the best single reactive element is found to be a condenser whose impedance is equal to the impedance level at the cutoff frequency. This gives a value of $K = 1 - j\nu$ which when substituted in (5) shows that the input impedance is to a first approximation a constant times the terminal impedance. To correct for the reactive component of the input impedance an inductance having an impedance jZ_1/ν which is equal to the input impedance level at cutoff is shunted across the input. The resulting impedance transforming network consists of an exponential line with a series capacitance at the high impedance end and a shunt inductance at the low impedance end. When terminated in a resistance load at either end equal to the impedance level at that end the input impedance, to a first approximation, is a resistance equal to the impedance level at the input end. In fact the deviations of the input impedance from the ideal for transmission in one direction are just the reciprocal of those for transmission in the other direction.

The magnitudes of the series capacitance and shunt inductance that give the improved network may be expressed in terms of the electrostatic capacitance and loop inductance of the line. Simple calculation shows that the required series capacitance is equal to $2/(k - 1)$ times the electrostatic capacitance of the line and the required shunt inductance is equal to the same factor times the total inductance of the line.

There is an interesting relationship between these terminations and a simple high-pass filter. The LC product of the shunt and series arms of the filter resonates at f_1. If an ideal transformer with transformation ratio k is inserted between the shunt inductance and the series capacitance, the capacitance becomes C/k and the new LC resonates at $f_1\sqrt{k}$. This is the same frequency at which the series capacitance and shunt inductance that are added to the terminations of the exponential line resonate. Furthermore the reactance of the shunt inductance is equal to the impedance level at the cutoff frequency and the reactance of the series capacitance is equal to the impedance level at the cutoff frequency exactly as in the case of the high-pass filter.

By using the exponential line it is possible to construct a network with properties that no network with lumped circuit elements possesses, namely, a high-pass impedance transforming filter.

Critical Lengths

Besides the characteristics of the exponential line that are substantially independent of the length of the line, it has properties that

depend on the length of the line that are analogous to those of a uniform line a half wave-length or quarter wave-length long. For non-dissipative lines above the cutoff frequency (5) becomes

$$Z_1 = \frac{K \cos\left(\frac{\eta}{2} - 2\xi\right) + j \sin\frac{\eta}{2}}{\cos\left(\frac{\eta}{2} + 2\xi\right) + jK \sin\frac{\eta}{2}} Z_1. \tag{8}$$

When the line is an integral number of half wave-lengths long ($2\eta = \pi$) this reduces to

$$Z_1 = KZ_1 = kZ_2, \tag{9}$$

which says that the input impedance is equal to the impedance transformation ratio times the load impedance. The length of exponential line that corresponds to a quarter-wave uniform line differs from an odd multiple of a quarter wave-length by an amount such that

$$\tan\left(\frac{\eta - (2n+1)\pi}{2}\right) = \frac{K^2 - 1}{K^2 + 1} \tan 2\xi, \tag{10}$$

for which (8) becomes

$$Z_1 = \frac{Z_1 Z_2}{Z_2}. \tag{11}$$

Similar expressions exist for the step-down line, but $1/K$ must be substituted for K in (10) for the length corresponding to the quarter-wave uniform line.

With Dissipation

An exponential line is an improvement over the uniform "iron wire" line as a resistance load that will dissipate a large amount of power.

Provided the attenuation is not too large the current and voltage distribution will be the same as for a non-dissipative line except for the additional power loss so that we may use the equations for an exponential line even though the distributed series resistance and shunt leakage do not vary exponentially with distance.

Suppose that the conductor size and resistance that will just dissipate the desired input power result in an attenuation constant α_0 for a uniform transmission line. To a first approximation the conductors can carry the same current irrespective of the impedance level. The current wave will be given by the first term of equation (2) which becomes

$$i = i_0 e^{-(\delta/2)x - \alpha_0 x},$$

except for a phase factor. In order that the current will not increase, $\delta = -2\alpha$. The actual attenuation "constant,"

$$\alpha_x \sim \left(\frac{R}{2Z_x} + \frac{GZ_x}{2}\right)\left(1 + \frac{1}{2}\frac{f_1^2}{f^2}\right) \sim \alpha_0 e^{2\alpha_0 x}\left(1 + \frac{1}{2}\frac{f_1^2}{f^2}\right), \quad (13)$$

will increase with distance down the line so that the current will decrease but not as rapidly as with a uniform line. The total attenuation in nepers is approximately

$$\left(1 + \frac{1}{2}\frac{f^2}{f_1^2}\right)\int_{x=0}^{l} \alpha_0 e^{2\alpha_0 x} dx = \left(1 + \frac{1}{2}\frac{f^2}{f_1^2}\right)\left(\frac{e^{2\alpha_0 l} - 1}{2}\right). \quad (14)$$

At the point where the attenuation of the uniform line is 6 db the tapered line has an additional attenuation of 7 db above the uniform line or a total attenuation of more than twice. The current has been reduced to less than half. Here an improvement may be made by increasing the dissipation by either changing the wire size or resistivity of the conductor. A greater improvement would result from changing the resistivity because then the capacity for heat dissipation would be the same. Suppose, however, that one conductor material is to be used throughout and the dissipation capacity is proportional to the wire surface; then at this point the wire size could be reduced to 1/2, doubling the attenuation factor. It is already 4 times that for the uniform line, so this increases it to 8 times. The resulting total attenuation is 30 db in a length that would have less than 7 db if the line were uniform. If this attenuation were required the length of line could be reduced by a factor of about 4.4. Of course the spacing is very close at the end of this line, but the line could be shorted at the end. This would approximately double the current at the end, but here again the current carrying capacity of the line is more than double the current traveling down the line. With the line shorted the reflected current would be 60 db down, which would not affect the input impedance appreciably. For the first 13 db of attenuation the impedance of the line would be relatively free from changes due to changes in spacing resulting from wind, etc. When the spacing is small enough to be affected by wind, vibration, etc., the attenuation will be great enough to suppress these small irregularities.

Experiment

In order to verify the foregoing theoretical development, measurements have been made on several experimental lines. Figure 4 shows the results of measurements on two such lines. These lines were

constructed of No. 12 tinned copper. At the low impedance end the strain was taken by a victron insulator which also served as a line spreader and terminal mounting. At the high impedance end the strain was taken by 1/4" manila rope without other insulation. The line spacing was adjusted by "lock stitch" tension insulators spaced 1 meter and 1/2 meter apart on the low and high impedance end respectively of the 9-meter line. The 3-meter line was supported at the 1/4, 1/2, 2/3, 3/4 and 7/8th points.

The impedance was measured by the substitution method. To facilitate the substitution of the reactive component of the line it was bridged by an antiresonant circuit. Pencil leads calibrated on direct current were used as the resistance standards. Type BW IRC 1/2

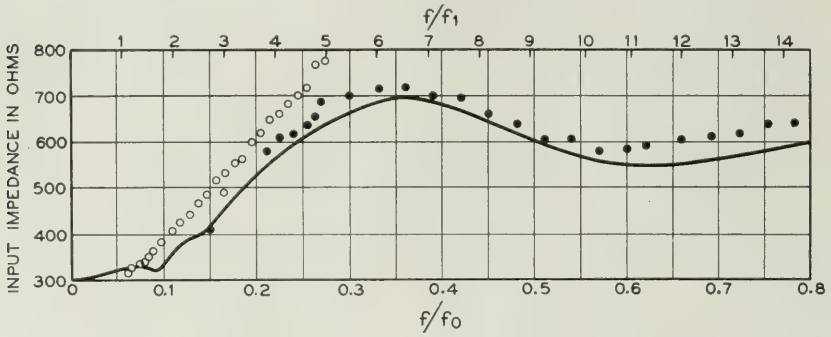


Fig. 4—Input impedance characteristic. Comparison of theoretical curve with experimental points for 600 : 300 ohm lines.
Solid circles—9-meter line.
Open circles—3-meter line.

watt resistances were used for terminations. The solid circles of Fig. 4 represent measurements on the 9-meter line. The agreement with theory is as good as is usually found for actual "uniform lines." In order to check the theory further toward the lower frequency end—beyond the range of the measuring equipment—measurements were made on a 3-meter line. These measurements are shown by the open circles. The agreement with theory is not so good, but here the lengths of the connecting leads are an appreciable fraction of the length of the line.

Preliminary tests on a full size model of exponential line impedance transformer showed deviations from the theoretical that might be attributed to improper termination, irregularities along the line, irregularities introduced at the change in conductor size or capacitance of the spacing insulators. Since it was impossible to determine which of these was the predominant cause of the deviations from the ideal, it

was decided to introduce each of these factors one at a time. This test was made on a 600 : 300 ohm line constructed of No. 6 copper wire with lockstitch insulators except at the terminals. The correct termination was obtained by tests on a uniform 300 ohm line with the same physical structure at the termination. Of necessity the tying of the wire to the strain insulators at the end introduced a shunt capacitance which augmented the inherent additional capacitance due the "end effect." This additional capacitance is equal to that of a short length of line.

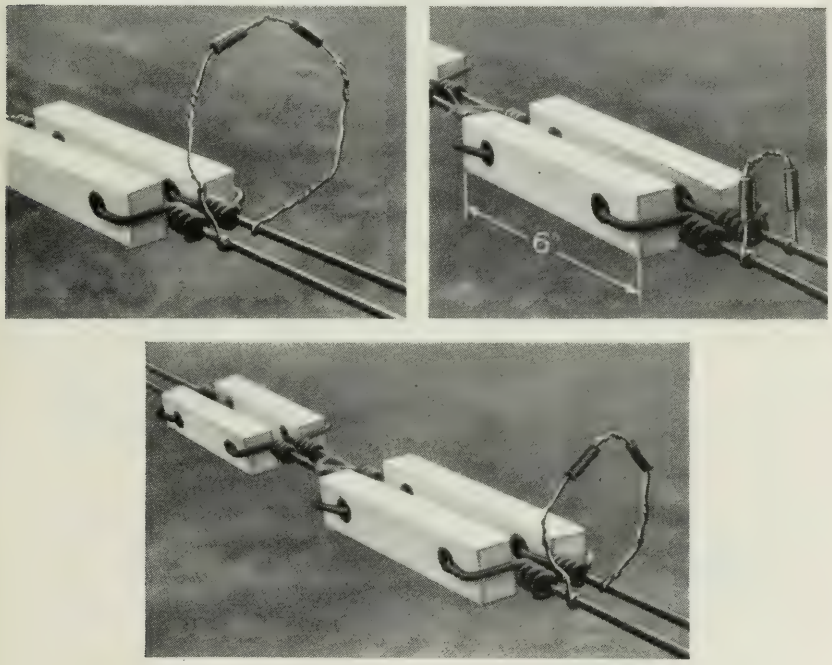


Fig. 5—Photographs of terminations of 300 ohm line.
upper right for curve 1
lower for curve 2
upper left for curve 3
} of Fig. 6.

If the correct amount of inductance is inserted in series with the resistance load the combined effect of the additional capacitance and inductance becomes the same as the addition of a small length of line for all frequencies up to those for which this length of line is an appreciable fraction of a wave-length. Accordingly a small amount of inductance was inserted in series with the resistance as shown in the right picture of Fig. 5. The input impedance of the uniform line with this termination is given by "Experimental Curve 1" at the bottom of Fig. 6.

A three-inch length of No. 18 wire was inserted as shown in the lower picture of Fig. 5 and "Experimental Curve 2" resulted. This reduced the irregularities in the input impedance to about half, so another three

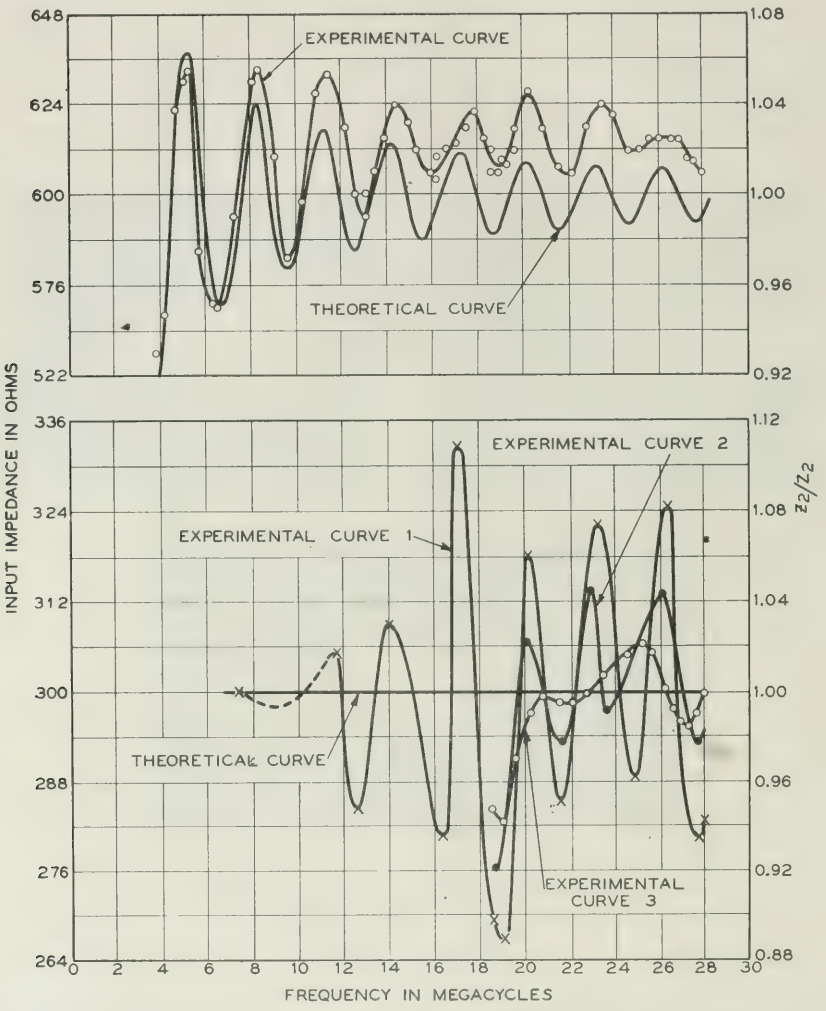


Fig. 6—Lower. Experimental input impedance characteristics of 300 ohm line with terminations shown in Fig. 5. Upper. Input impedance characteristics of 50-meter 600 : 300 ohm line of No. 6 conductors.

inches were inserted, resulting in "Experimental Curve 3." Here the maxima and minima are displaced, indicating that the effect of the stray capacitance has been reduced to the same order of magnitude as that due to the deviation of the resistance from the desired value. This

termination was accordingly removed to the exponential line, resulting in the "Experimental Curve" at the top of Fig. 6. It agrees within experimental error with the "Theoretical Curve." The slight vertical displacement of the experimental curve at the higher frequencies is attributed to deviations in the impedance of the pencil lead, which was

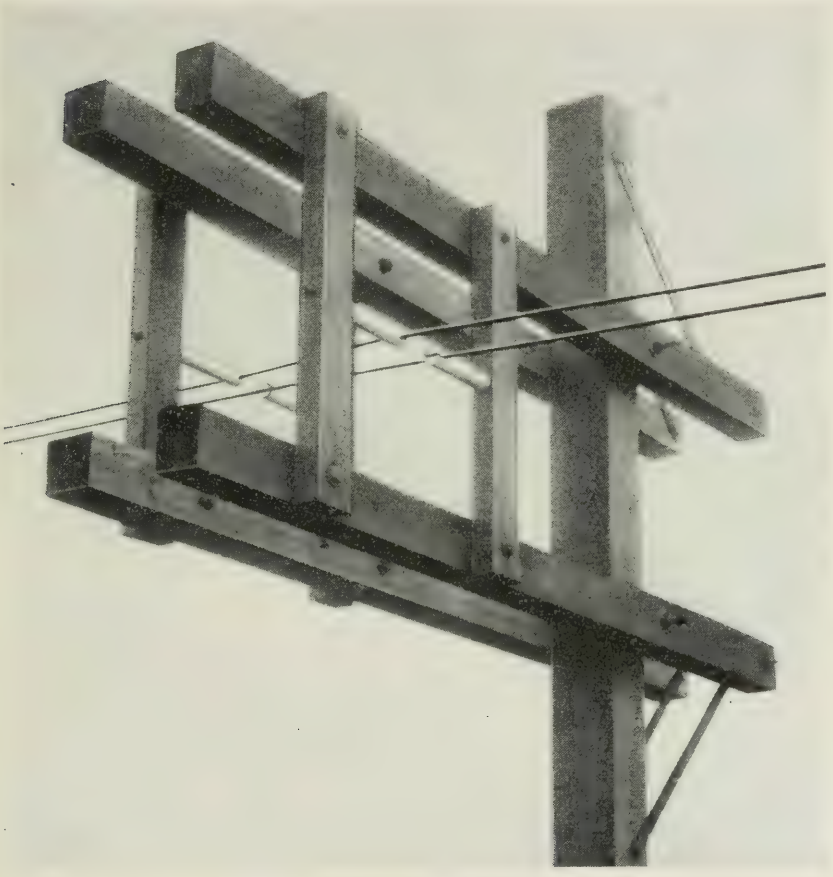


Fig. 7—Photograph of one of the changes in conductor size.

used as a resistance standard, from a pure resistance equal to its direct current value.

To increase the power carrying capacity of the exponential line, one was built with larger wire size at the lower impedance end. This increased the breakdown voltage by increasing the spacing and conductor diameter and at the same time increased the current carrying capacity by decreasing the resistance and increasing the heat dissipat-

ing capacity of the conductors. This was a 600 : 300 ohm line constructed of 20 meters No. 6 wire, 10 meters 1/4″ tubing and 20 meters 3/8″ tubing. Here again the correct termination was determined by measurements on a 300 ohm uniform line of 3/8″ tubing. The total length of terminating loop that gave the best termination was 6½″ in this case compared with 10½″ for the 300 ohm line of No. 6 wire. Since no attempt was made to reduce the variations in input impedance to less than ± 1 per cent these lengths may be as much as an inch off.

These measurements indicated that the exponential line would perform satisfactorily as an impedance transformer if it could be constructed to have the desired mechanical features without impairing its electrical properties. The greatest difficulty appeared to reside in the

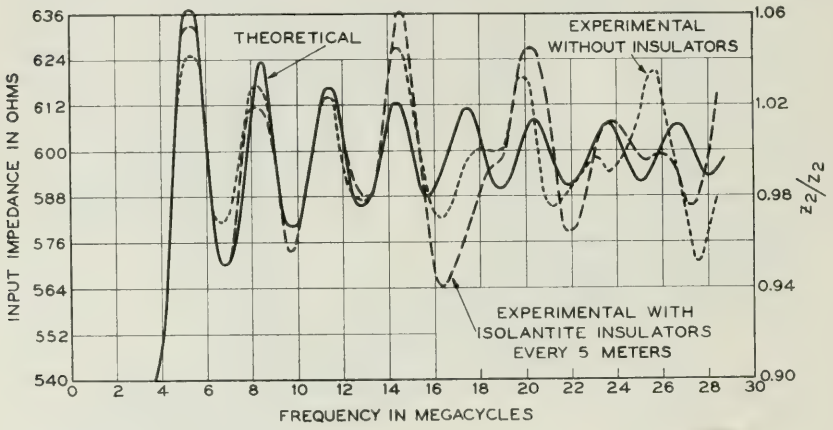


Fig. 8—Input impedance characteristics of 50-meter 600 : 300 ohm line of 3/8″, 1/4″ and No. 6 conductors.

insulators. Special isolantite insulators were designed that would be satisfactory commercially and still keep the additional capacity to a reasonable value. Figure 7 shows the construction of the line at the supporting poles where the conductor size changes.

The results of measurements on this line are shown in Fig. 8. The solid curve was calculated from the equations developed earlier. The two broken curves are the results of measurements on the line, one without insulators and one with insulators. While the insulators affect the line somewhat they do not increase the deviation from the ideal appreciably. [The improvement in the agreement between experiment and theory in this set of curves over that in Fig. 4 is presumably due to the fact that the comparison resistance for Fig. 8 consisted of 3-IRC

resistances instead of the pencil lead. With the fixed IRC resistance it was, of course, impossible to adjust the standard to exactly the same value as the unknown. In this case the small difference was determined by using the slope of the rectifier voltmeter calibration.] This line has a maximum deviation from the desired input impedance of ± 6 percent for all frequencies above 4.2 mc. (Measurements were made up to 28 mc.) The phase angle of the input impedance was found to be zero

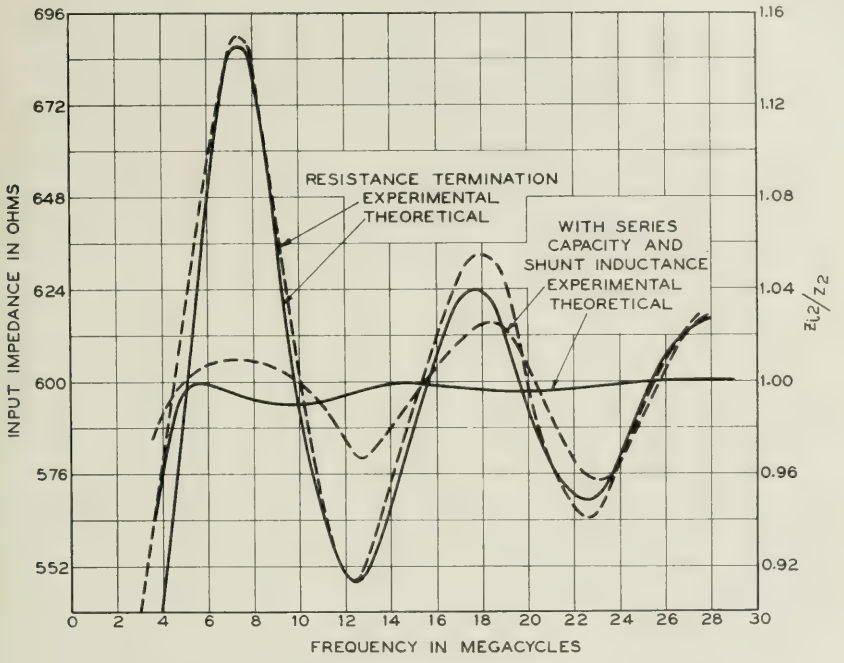


Fig. 9—Input impedance characteristics of 15-meter 600 : 300 ohm line of No. 6 conductors

within the accuracy of measurement. From theory the phase angle would be expected to vary between − 0° and + 3°.

The curves of Fig. 9 refer to a 600 : 300 ohm line of No. 6 wire 15 meters long. With resistance termination this line has rather large variations in the input impedance but with the addition of the proper reactances the input impedance is flatter than the longer line with resistance termination. At the lower frequencies where the variations in the input impedance were large without the reactive networks, their addition gives approximately the expected improvement. At the higher frequencies the inductance was approximately anti-resonated

by its distributed capacity and the input impedance approaches that for the resistance termination.

Conclusion

Theory indicates that the exponential line may be used as an impedance transformer over a wide frequency range. The results of experiment show that the desired characteristic can be realized in practice. Among the applications of the exponential line may be mentioned its use in transforming the impedance level back to its original value after the paralleling of two transmission lines feeding two antennas. It could be used to transform the input impedance of a rhombic antenna down to the usual 600-ohm level of open wire transmission lines. If twin coaxial lines are used inside the transmitter building to eliminate undesired feedback, coupling, etc., the exponential line could be used to transform from the highest practical impedance level of such lines to a practical level of the more economical open wire lines for use outside the building.

Appendix

The exponential line is a non-uniform line so that the terms "characteristic impedance" and "surge impedance" of an exponential line are not synonymous. The terms "surge impedance" [1] and "nominal characteristic impedance" [2] may be used synonymously for the characteristic impedance of the uniform line that has the same distributed constants as the non-uniform line at the point in question. Expressed as functions of the distributed "constants" of the line they are the square root of the ratio of the distributed series impedance to the distributed shunt admittance at the point along the line in question. It will be expedient to refer to the nominal characteristic impedance as the impedance level at the point in question. Schelkunoff [3] has defined the characteristic impedances as the ratio of voltage to current at the point in question for each of the two traveling waves of which

[1] The term "surge impedance" is defined by A. E. Kennelly on page 73 of "The Applications of Hyperbolic Functions to Electrical Engineering Problems" (McGraw-Hill 1916) as follows: "The surge impedance of the line is not only the natural impedance which it offers everywhere to surges of the frequency considered, but it is also the initial impedance of the line at the sending end." Hence the "surge impedance" should be independent of the configuration of the line except at the point in question and in particular it should be equal to that for a uniform line constructed so as to have the same dimensions everywhere as the non-uniform line has at the point in question.

[2] The word nominal as used here has the same meaning as in "nominal iterative impedance" as used by K. S. Johnson in "transmission circuits for telephone communication" (Van Nostrand 1925).

[3] S. A. Schelkunoff, "The Impedance Concept and its Application to Problems of Reflection, Refraction, Shielding and Power Absorption," *Bell System Technical Journal*, *17*, 17–48, January, 1938.

the steady state condition is composed. At each point an exponential line has two characteristic impedances which are different and depend upon the frequency as well as the position along the line.

Because of the change of impedance level, the propagation constants for the voltage and current differ, so that it is convenient to consider the transfer constant [4] which may be defined as half the sum of the voltage and current propagation constants.

[4] Compare with the definition of "image transfer constant" as given by K. S. Johnson in "Transmission Circuits for Telephone Communication."

The Bridge Stabilized Oscillator*

By L. A. MEACHAM

A new type of constant frequency oscillator of very high stability is presented. The frequency controlling resonant element is used as one arm of a Wheatstone resistance bridge. Kept in balance automatically by a thermally controlled arm, this bridge provides constancy of output amplitude, purity of wave form, and stabilization against fluctuations in power supply or changes in circuit elements. A simple one-tube circuit has operated consistently with no short-time frequency variations greater than ± 2 parts in 10^8. Convenient means are provided for making precision adjustments over a narrow range of frequencies to compensate for long-time aging effects.

Description of the circuit is followed by a brief linear analysis and an account of experimental results. Operating records are given for a 100 kc. oscillator.

Introduction

THE problem of improving the stability of constant frequency oscillators may be divided conveniently into two parts, one relating to the frequency controlling resonant element or circuit, and the other to the means for supplying energy to sustain oscillations. The ideal control element would be a high-Q electrical resonant circuit, or a mechanical resonator such as a tuning fork or crystal, whose properties were exactly constant, unaffected by atmospheric conditions, jar, amplitude of oscillation, age, or any other possible parameter. The ideal driving circuit would take full advantage of the resonator's constancy by causing it to oscillate at a stable amplitude and at a frequency determined completely by the resonator itself, regardless of power supply variations, aging of vacuum tubes or other circuit elements, or the changing of any other operating condition.

This paper, concerning itself principally with the second part of the problem, describes an oscillator circuit which attains a very close approximation to the latter objective. The "Bridge Stabilized Oscillator" provides both frequency and amplitude stabilization, and as it operates with no tube overloading, it has the added merit of delivering a very pure sinusoidal output.

Oscillator Circuit

The bridge stabilized oscillator circuit, shown schematically in Fig. 1, consists of an amplifier and a Wheatstone bridge. The amplifier out-

* Presented at Thirteenth Annual Convention of Institute of Radio Engineers, New York City, June 16, 1938. Published in *Proc. I. R. E.*, October 1938.

put is impressed across one of the diagonals of the bridge, and the unbalance potential, appearing across the conjugate diagonal, is applied to the amplifier input terminals. One of the four bridge arms, R_1, is a thermally controlled resistance; two others, R_2 and R_3, are fixed resistances, and the fourth, $Z_4 = R_4 + jX_4$, is the frequency-controlling resonant element.

In this discussion Z_4 is assumed to represent a crystal suitable for operation at its low-impedance or series resonance. A coil and condenser in series could be substituted, and even a parallel-resonant control element might be used by exchanging its position in the bridge

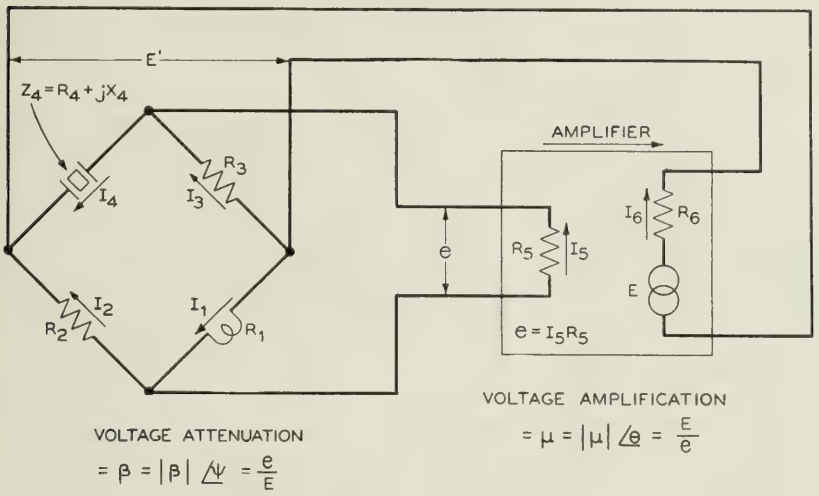


Fig. 1—Schematic circuit diagram of bridge stabilized oscillator.

with R_2 or R_3. Operating a crystal at series resonance has the advantage of minimizing effects of stray capacitance.

The bridge is kept as nearly in exact balance as possible. Assuming that R_1, R_2 and R_3 are pure resistances, we may write for exact reactive balance,

$$X_4 = 0,$$

and for exact resistive balance,

$$\frac{R_1}{R_2} = \frac{R_3}{R_4}.$$

In order that the circuit may oscillate, a slight unbalance is required. Accordingly R_1 must be given a value slightly smaller than $(R_2R_3)/R_4$,

so that the attenuation through the bridge is just equal to the gain of the amplifier.

It is evident that if all the bridge arms had fixed values of resistance, the attenuation of the bridge would be very critical with slight changes in any arm. This would obviously be undesirable, for the circuit would either fail to oscillate, or else build up in amplitude until tube overloading occurred. The thermally controlled resistance R_1 eliminates this difficulty. This arm has a large positive temperature coefficient of resistance, and is so designed that the portion of the amplifier output which reaches it in the bridge circuit is great enough to raise its temperature and increase its resistance materially. A small tungsten-filament lamp of low wattage rating has been found suitable. It functions as follows:

When battery is first applied to the amplifier, the lamp R_1 is cold and its resistance is considerably smaller than the balance value. Thus the attenuation of the bridge is relatively small, and oscillation builds up rapidly. As the lamp filament warms, its resistance approaches the value for which the loss through the bridge equals the gain of the amplifier. If for some reason R_1 acquires too large a resistance, the unbalance potential e becomes too small or possibly even inverted in phase, so that the amplitude decreases until the proper equilibrium is reached.

This automatic adjustment stabilizes the amplitude, for the amount of power needed to give R_1 a value closely approaching $(R_2R_3)/R_4$ is always very nearly the same. A change in the amplifier gain would cause a readjustment of the bridge balance, but the resulting variation in R_1 or in the amplifier output would be extremely small. The operating temperature of the lamp filament is made high enough so that variations in the ambient temperature do not affect the adjustment appreciably.

No overloading occurs in the amplifier, which operates on a strictly Class A basis, nor is any non-linearity necessary in the system other than the thermal non-linearity of R_1. As the lamp resistance does not vary appreciably during a high-frequency cycle, it is not a source of harmonics (or of their intermodulation, which Llewellyn [1] has shown to be one of the factors contributing to frequency instability).

In contrast to the lamp, an ordinary non-linear resistance, of copper oxide for example, would not be suitable for R_1. A resistance of the thermally-controlled type having a negative temperature coefficient

[1] "Constant Frequency Oscillators," F. B. Llewellyn, *Proc. I. R. E.*, December 1931.

could be used by merely exchanging its position in the bridge with R_2 or R_3.

The frequency control exerted by the crystal depends upon the fact that the phase shift of the amplifier must be equal and opposite to that through the bridge. In the notation of Black,[2] applied to the circuit of Fig. 1,

$$\mu = \frac{E}{e} = |\mu| \underline{|\theta},$$

and

$$\beta = \frac{e}{E} = |\beta| \underline{|\psi}.$$

The condition for oscillation is

$$\mu\beta = 1 \underline{|0},$$

which implies that $|\mu\beta| = 1$ and $\theta = -\psi$.

The vector diagrams of Fig. 2 illustrate the frequency-stabilizing action of the bridge by showing the voltage relations therein for two values of amplifier phase shift, θ. When θ is zero, as in diagram A, the unbalance vector e is in phase with the generated voltage E applied to the bridge input, and thus all the vectors shown are parallel. They are displaced vertically from each other merely to clarify the drawing. The crystal is here constrained to operate at exact resonance.

In diagram B, the amplifier is assumed to have changed its phase for some reason by an amount far in excess of what would be anticipated in practice, θ here having a value of + 45 degrees. The important point to be observed is that the ratio of θ to the resulting change in the phase angle ϕ of the crystal impedance Z_4 is very large. That is, the crystal is still operating close to resonance in spite of the exaggerated change in the driving circuit. If the gain of the amplifier were greater, the action of the thermally controlled resistance would keep the amplifier output vector E practically the same in length, making the unbalance vector e correspondingly shorter. The angle ϕ would therefore have to be more acute for the same value of θ, and it follows that with increased gain the crystal is held closer to true resonance and the stability is improved.

When θ equals zero, changes in $|\mu|$ do not affect the crystal operating phase, but for any other small value of θ, gain variations cause slight readjustments of the angles between vectors. The amplifier should accordingly be designed for zero phase shift, and also, of course, should have as much phase stability as possible.

[2] "Stabilized Feedback Amplifiers," H. S. Black, *Bell System Technical Journal*, January 1934.

In this discussion the input and output impedances of the amplifier, R_5 and R_6, are assumed to be constant pure resistances. Actually, changes in the tube parameters or in certain circuit elements are likely to affect both the magnitude and the phase of these impedances. It may be shown, however, that such changes have the same effect upon the bridge and upon the frequency as do changes of about the same

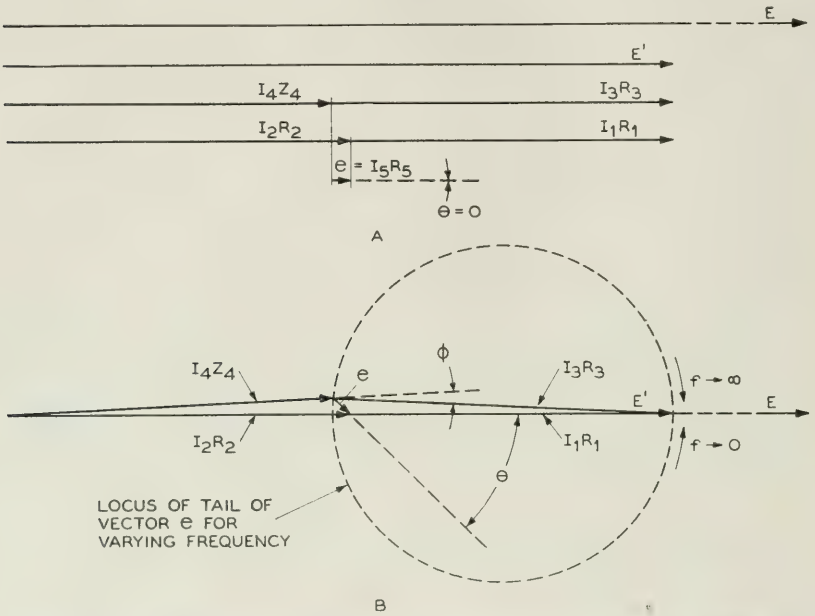


Fig. 2—Vector diagrams illustrating operation of bridge oscillator, with simplifying assumptions that R_5 is large and that E and E' are strictly in phase.

A—At resonance
$Z_4 = R_4 + j0$
$\theta = 0$
$R_1 < R_2 = R_3 = R_4$
B—Above resonance
$Z_4 = R_4 + jX_4$
X_4 Inductive
$\theta = +45°$
$R_1 < R_2 = R_3 = R_4 \ll R_5$

percentage in $|\mu|$ or θ; therefore all variations in the driving circuit external to the bridge may be assumed for convenience to be represented by variations in its gain and phase.

This leniency with regard to R_5 and R_6 does not apply to the other bridge resistances, however. R_1, R_2 and R_3 are directly responsible for the crystal's operating phase and amplitude; they should be made as stable and as free from stray reactance as possible.

The effect of the bridge upon harmonics of the oscillation frequency is of interest. Harmonics, being far from the resonant frequency of the crystal, are passed through the bridge with little attenuation but with a phase reversal approximating 180 degrees, as illustrated by the dotted locus in Fig. 2. Thus if the amplifier were designed to cover a band broad enough to include one or more harmonics and if care were taken to avoid singing at some unwanted frequency, a considerable amount of negative feedback could be applied to the suppression of the harmonics in question.

Circuit Analysis

In the following section, expressions are derived for the frequency of oscillation in terms of the gain and phase shift of the amplifier, the Q of the crystal, and values of the bridge resistances. It is assumed that the latter are constant and non-reactive, and therefore, as explained previously, that all sources of frequency fluctuations apart from changes in the crystal itself appear as variations in $|\mu|$ or θ. Because the bridge oscillator does not rely upon non-linearity in the ordinary sense to limit its amplitude, the analysis can be based reasonably on simple linear theory.

In the near vicinity of series resonance the crystal may be represented accurately by a resistance R_4, inductance L and capacitance C, connected in series. The reactive component of the crystal's impedance is accordingly

$$X_4 = \omega L - \frac{1}{\omega C} = \frac{\omega^2 LC - 1}{\omega C}. \tag{1}$$

Solving for the frequency,

$$\begin{aligned}\omega &= \frac{X_4}{2L} + \sqrt{\left(\frac{X_4}{2L}\right)^2 + \frac{1}{LC}} \\ &= \frac{1}{\sqrt{LC}}\left[\frac{X_4}{2}\sqrt{\frac{C}{L}} + \sqrt{1 + \left(\frac{X_4}{2}\sqrt{\frac{C}{L}}\right)^2}\right] \\ &= \frac{1}{\sqrt{LC}}\left[1 + \frac{X_4}{2}\sqrt{\frac{C}{L}} + \frac{1}{2}\left(\frac{X_4}{2}\sqrt{\frac{C}{L}}\right)^2 \right. \\ &\qquad \left. - \frac{1}{2}\cdot\frac{1}{4}\left(\frac{X_4}{2}\sqrt{\frac{C}{L}}\right)^4 + \cdots\right]. \end{aligned} \tag{2}$$

Near series resonance, $(X_4/2)\sqrt{(C/L)} << 1$. We therefore disregard powers higher than the first in the series expansion above and obtain the close approximation,

$$\omega \doteq \frac{1}{\sqrt{LC}}\left[1 + \frac{X_4}{2}\sqrt{\frac{C}{L}}\right]. \tag{3}$$

The frequency deviation from resonance, expressed as a fraction of the resonant frequency f_0, is thus

$$\frac{f - f_0}{f_0} = \frac{\omega - \omega_0}{\omega_0} \doteq \frac{X_4}{2}\sqrt{\frac{C}{L}}, \tag{4}$$

and in the region of interest, where ωL and $1/\omega C$ are approximately equal,

$$\frac{f - f_0}{f_0} \doteq \frac{X_4}{2\omega L} = \frac{X_4}{2QR_4}. \tag{5}$$

Considering now the bridge circuit, and applying well-known equations,[3] we obtain

$$\beta = \frac{I_5 R_5}{E} = \frac{AR_4 - jBX_4}{MR_4 + jNX_4}, \tag{6}$$

in which

$$\left.\begin{aligned} A &= R_5(R_2R_3 - R_1R_4), \\ B &= R_1R_4R_5, \\ M &= (R_1 + R_2)(R_3R_4 + R_5R_6) + (R_3 + R_4)(R_1R_2 + R_5R_6) \\ &\quad + (R_5 + R_6)(R_1R_4 + R_2R_3) + R_5(R_1R_3 + R_2R_4) \\ &\quad + R_6(R_1R_2 + R_3R_4), \\ \text{and} \\ N &= R_4(R_1 + R_3 + R_5)(R_2 + R_6) + R_1R_4(R_3 + R_5). \end{aligned}\right\} \tag{7}$$

The condition for oscillation, as mentioned previously, is $\mu\beta = 1\underline{|}0$. Putting $\mu = \mu_1 + j\mu_2$, we may write

$$(\mu_1 + j\mu_2)\left(\frac{AR_4 - jBX_4}{MR_4 + jNX_4}\right) = 1, \tag{8}$$

which gives the pair of equations

$$\mu_1 AR_4 + \mu_2 BX_4 - MR_4 = 0 \tag{9}$$

and

$$\mu_2 AR_4 - (\mu_1 B + N)X_4 = 0. \tag{10}$$

For the special case in which the amplifier phase shift is zero ($\mu_2 = 0$), these become

$$\mu_1 = \frac{M}{A} = |\mu| \tag{11}$$

and

$$X_4 = 0. \tag{12}$$

[3] "Transmission Circuits for Telephonic Communication," K. S. Johnson, pp. 284–5. D. Van Nostrand Company.

The latter equation indicates that the frequency is then independent of changes in any of the circuit parameters except the crystal, which must operate exactly at resonance.

If the phase of μ differs only slightly from zero, so that μ_2 is very small, then it may be inferred from continuity considerations that the frequency is still very nearly independent of all circuit parameters, except of course variations in θ, the phase of μ. When θ is limited to values for which $\mu_2 B X_4 << \mu_1 A R_4$, (11) still applies closely. Substitution into (10) gives

$$X_4 \doteq \frac{MR_4}{B\mu_1 + N} \cdot \frac{\mu_2}{\mu_1} \doteq \frac{MR_4\theta}{B|\mu| + N}, \tag{13}$$

and finally from (5) and (13) we obtain the frequency deviation in the form

$$\frac{f - f_0}{f_0} \doteq \frac{M\theta}{2Q(B|\mu| + N)}. \tag{14}$$

As noted above, this expression applies accurately only when θ is small, as it should be in a well designed bridge oscillator.

The effect of variations in the amplifier may be examined by differentiating (14). For changes in θ only,

$$\frac{df}{f_0}\Big]_\theta \doteq \frac{M}{2Q(B|\mu| + N)}\, d\theta, \tag{15}$$

and for those of $|\mu|$,

$$\frac{df}{f_0}\Big]_{|\mu|} \doteq -\frac{BM\theta}{2Q(B|\mu| + N)^2}\, d|\mu|. \tag{16}$$

Equations (15) and (16) have been found to be closely in accord with experiment, although the differentiation is not rigorously allowable (B, M and N being only approximately constant).

In the special case where all the fixed bridge resistances (R_2 to R_6 inclusive) are equal, and $|\mu|$ is large enough so that R_1 has nearly the same value, (14), (15) and (16) reduce to the following:

$$\frac{f - f_0}{f_0} \doteq \frac{8\theta}{Q(|\mu| + 8)}, \tag{17}$$

$$\frac{df}{f_0}\Big]_\theta \doteq \frac{8}{Q(|\mu| + 8)}\, d\theta, \tag{18}$$

$$\frac{df}{f_0}\Big]_{|\mu|} \doteq -\frac{8\theta}{Q(|\mu| + 8)^2}\, d|\mu|. \tag{19}$$

These expressions show, as did the vector diagrams, that for optimum stability the amplifier phase shift should be made approximately zero, the crystal should have as large a value of Q (as low a decrement) as possible, and the amplifier should have high gain. Linearity in the amplifier is also desirable, to minimize the modulation effects described by Llewellyn.[1] When present, these effects appear as variations in $|\mu|$ and θ.

One of the significant differences between the bridge oscillator and other oscillator circuits is the fact that its frequency stability is roughly proportional to $|\mu|$. This relationship holds at least for amounts of gain that can be dealt with conveniently. Although increased gain is generally accompanied by larger variations in phase, the two are not necessarily proportional. For example, if greater stability were required for some precision application than could be achieved with a single-tube bridge oscillator, and if the constancy of the crystal itself warranted further circuit stabilization, it could be obtained by adding another stage. The phase fluctuations in the amplifier might possibly be doubled, but the value of $|\mu|$ would be multiplied by the amplification of the added tube, giving an overall improvement.

To illustrate the high order of stability provided by a bridge oscillator, let us consider a model composed of a single-tube amplifier and a bridge in which all the fixed resistances are made equal to that of the crystal. We will assume the crystal to have a reasonably high [4] Q of 100,000. The amplifier phase, let us say, is normally zero, but may possibly vary $\pm$ 0.1 radian ($\pm$ 6 degrees), and the value of $|\mu|$, nominally 400, may change $\pm$ 10 per cent. From (18) and (19) we find

$$\left.\frac{\Delta f}{f_0}\right]_{\theta} = \pm \frac{(8)(0.1)}{(100{,}000)(360 + 8)} = \pm 2.17 \times 10^{-8},$$

and (when θ has its maximum value of 0.1 radian)

$$\left.\frac{\Delta f}{f_0}\right]_{|\mu|} = \pm \frac{(8)(0.1)(40)}{(100{,}000)(360 + 8)^2} = \pm 2.36 \times 10^{-9}.$$

This example represents the degree of stabilization against circuit fluctuations that can be obtained with a simple form of the oscillator operating under poorly controlled conditions. By stabilizing the power supply and other factors affecting $|\mu|$ and θ, and by increasing the gain, the frequency variations arising in the driving circuit may be made negligible compared to the variations found at present in the properties even of the best mounted crystals.

[4] For crystals in vacuum, values of Q as great as 300,000 have been obtained.

Experiment

The circuit diagram of an experimental bridge stabilized oscillator is shown in Fig. 3, and its photograph in Fig. 4. The amplifier unit consists of a single high-mu tube V_1 with tuned input and output transformers T_1 and T_2 and the usual power supply and biasing arrangements. The crystal, mounted in the cylindrical container at the left end of the panel, is one having a very low temperature coefficient of frequency at ordinary ambient temperatures. In Fig. 4 it is shown without provisions for temperature control. A high Q is obtained by clamping the crystal firmly at the center of its aluminum-

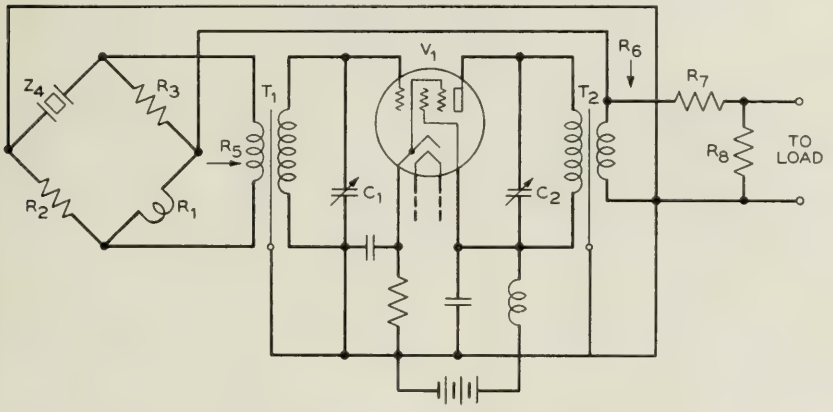


Fig. 3—Circuit of experimental bridge oscillator.

coated major faces between small metal electrodes ground to fit, and by evacuating the container.

Some of the circuit parameters are listed below:

R_1 = Tungsten-filament lamp,
R_2 = 100 ohms,
R_3 = 150 ohms,
Z_4 = 100 kc. crystal.
Characteristics at resonance:
R_4 = 114 ohms,
$X_L = X_C$ = 11,900,000 ohms,
Q = 104,000,
$R_5 = R_6$ = 150 ohms (approx.),
R_7 = 500 ohms,
R_8 = 200 ohms,
$|\mu|$ = 422 (52.5 *db* voltage gain from e to E).

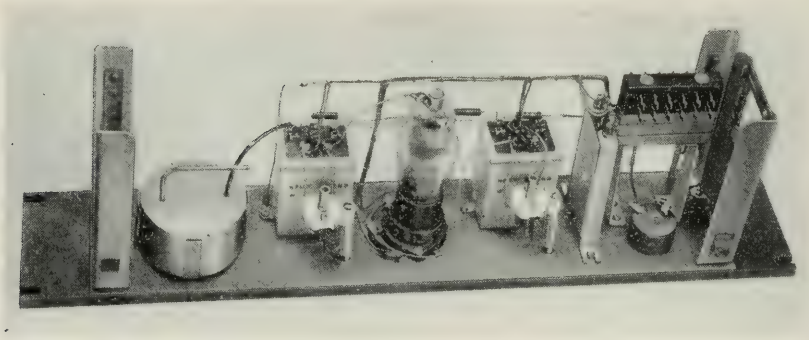


Fig. 4—Experimental bridge stabilized oscillator without provision for temperature control.

Figure 5 shows the resistance of the lamp R_1 plotted against the power dissipated in its filament. The large rise in resistance for small amounts of power is due to the effective thermal insulation provided by the vacuum surrounding the filament and to low heat loss by radiation. The lamp operates at temperatures below its glow point, assuring an extremely long life for the filament.

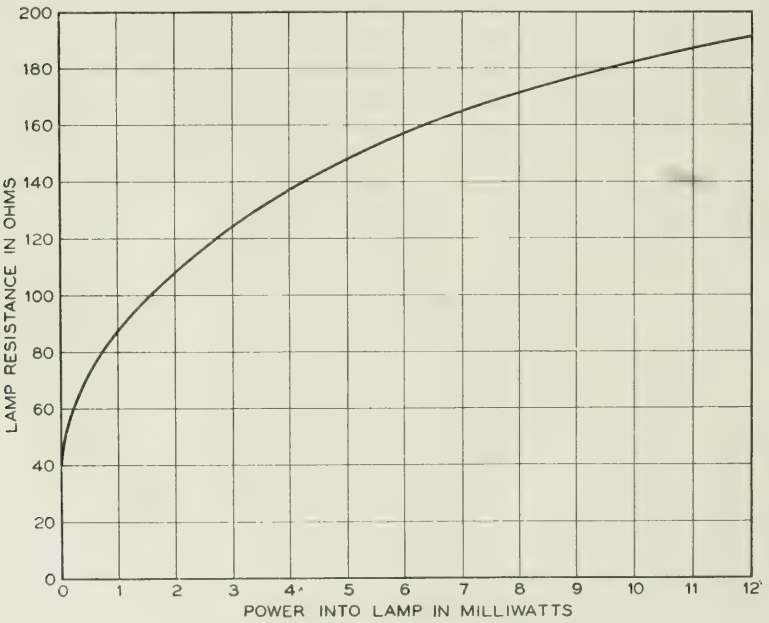


Fig. 5—Characteristic of lamp used for R_1.

The particular value assumed by R_1 in the circuit of Fig. 3 is approximately $(R_2R_3)/R_4 = [(100)(150)]/114 = 131.6$ ohms, and hence from Fig. 5 it follows that the power supplied to the lamp is about 3.7 milliwatts. The r.m.s. voltage across the lamp is computed to be 0.70 volt, and across the entire bridge, 1.23 volt. The power supplied to a load of 150 ohms through the pad composed of R_7 and R_8 is accordingly 0.22 milliwatt, or 6.6 *db* below 1 milliwatt, which is in agreement with measurements shown in Figs. 8 and 9, described below. These quantities are given to illustrate the fact that currents and voltages in

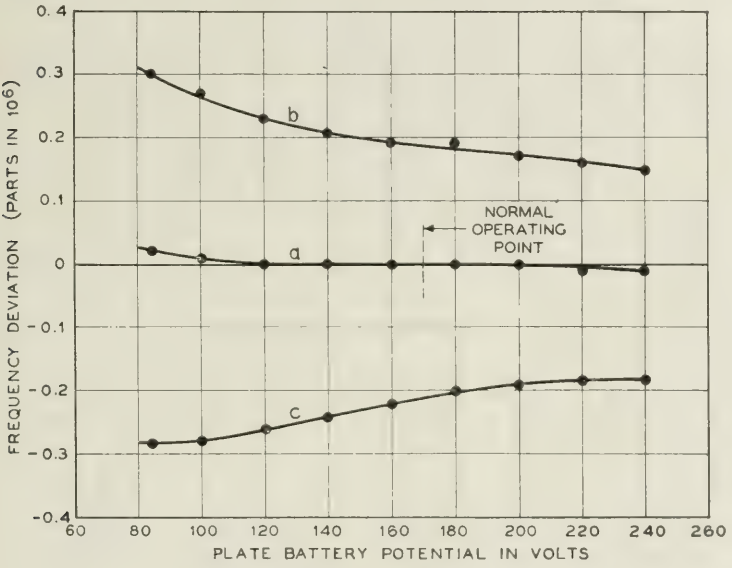


Fig. 6—Oscillator frequency vs. plate battery potential.

a—C_1 and C_2 tuned for maximum amplifier gain.
b—C_1 and C_2 decreased 5%.
c—C_1 and C_2 increased 5%.

this type of oscillator may be calculated readily from the values of the circuit elements, and without reference to the power supply voltages or the tube characteristics except to assume that they give the amplifier sufficient gain to operate the bridge near balance, and that tube overloading does not occur at the operating level.

Experimental performance curves for the circuit of Fig. 3 are presented in Figs. 6 to 11 inclusive. Figure 6 shows frequency deviation plotted against plate battery voltage for several settings of the grid and plate tuning condensers. For curve *a* the amplifier was adjusted at maximum gain, corresponding approximately to zero phase shift as

well. Here the frequency varied not more than one part in one hundred million for a voltage range from 120 to 240 volts. Curve b was taken with the two tuning capacitances C_1 and C_2 decreased 5 per cent from their optimum settings, and curve c with both capacitances increased 5 per cent. These detunings introduced phase shifts of about $\pm$ 40° ($\pm$ 0.70 radian), decreased $|\mu|$ by 0.8 db and changed the frequency, as shown in Fig. 6, approximately $\pm$ 2 parts in ten million. Although the analysis should not be expected to apply

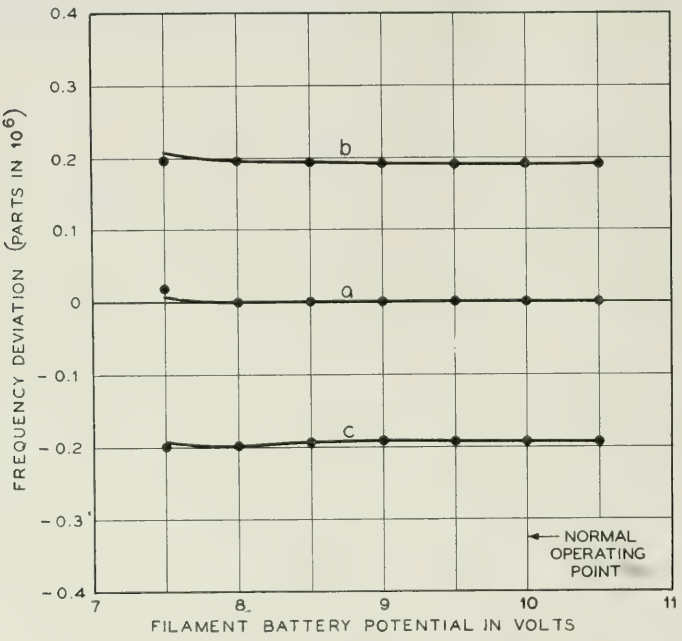


Fig. 7—Oscillator frequency vs. filament battery potential.
a—C_1 and C_2 tuned for maximum amplifier gain.
b—C_1 and C_2 decreased 5%.
c—C_1 and C_2 increased 5%.

accurately for such large phase shifts, calculation of the frequency deviations by means of (18) gives $\pm$ 1.4 parts in ten million—in fair agreement with the experimental results. As might be expected, curves b and c show somewhat less stability with battery voltage changes than does curve a.

Figure 7 presents a similar set of curves for variations of filament voltage. Here, for the "maximum-gain" tuning adjustment, a drop from 10 volts, the normal value, to 8 volts caused less than one part in one hundred million change of frequency, as shown in curve a.

In Fig. 8, the gain of the amplifier and the output level of the oscillator are plotted against plate battery voltage, while in Fig. 9 the same quantities are related to filament potential. These curves show that although power supply variations change the amplifier gain, they have but slight effect upon the amplitude of oscillation. This stabilization is produced, as explained heretofore, by the action of the lamp.

The oscillator was designed to work into a load of 150 ohms, its output impedance approximately matching this value. It might be expected that variations in the magnitude or phase angle of the load

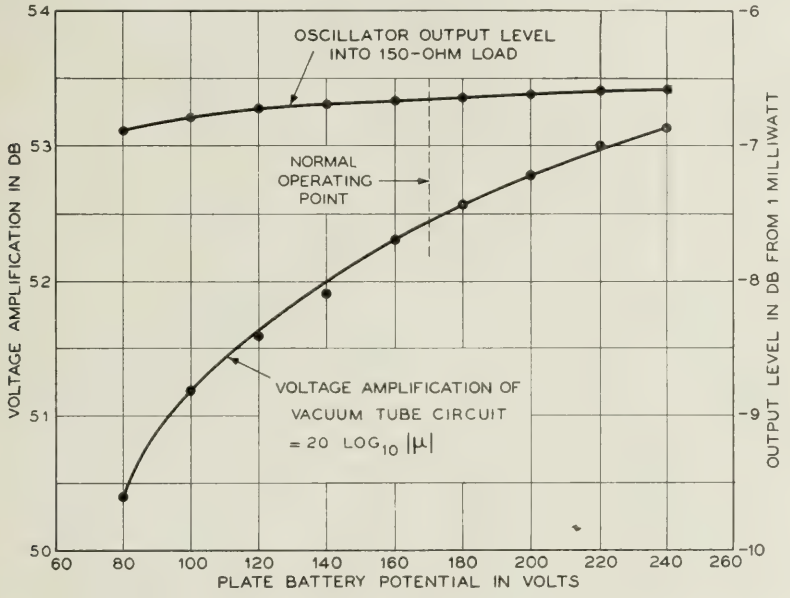


Fig. 8—Amplifier gain and oscillator output level vs. plate battery potential.

would affect the frequency materially even though a certain amount of isolation is provided by R_7 and R_8. However, measurements made with (1) a series of load impedances having a constant absolute magnitude of 150 ohms but with phase angles varying between $-90°$ and $+90°$ and (2) a series of resistive loads varying between 30 ohms and open circuit, showed less than one part in a hundred million frequency variation. Graphs of these results have not been included, since they practically coincide with the axis of zero frequency deviation.

The tuned transformers T_1 and T_2 in this experimental model precluded the suppression of harmonics by negative feedback, $|\mu|$ being small at the harmonic frequencies. The tuning itself provided suppression, however, so that the measured levels of the second and third

harmonics in the output current were respectively 67 db and 80 db below that of the fundamental. This purity of wave form is of course largely dependent upon the absence of overloading.

To correct any small initial frequency error of the crystal and to allow for subsequent aging, a small reactance connected directly in series with the crystal provides a convenient means of adjusting the frequency as precisely as it is known. This added reactance may be considered as modifying either of the reactances in the equivalent series resonant circuit of the crystal. Figure 10 shows that for a small

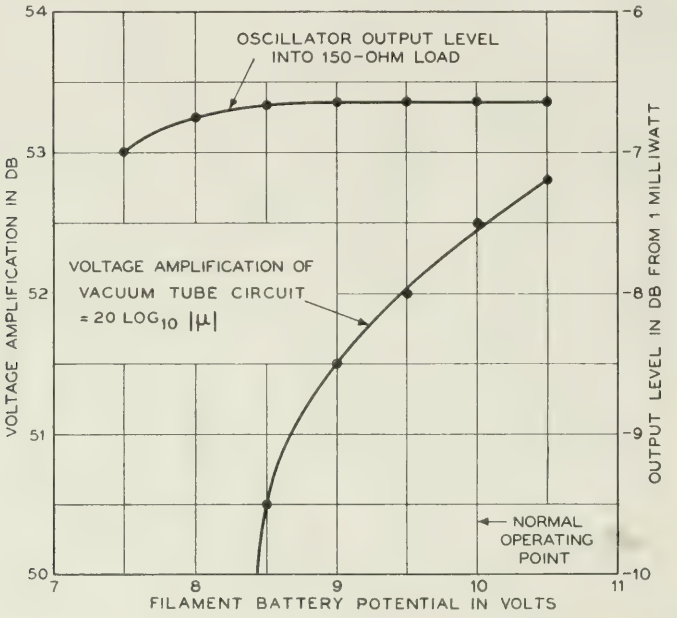


Fig. 9—Amplifier gain and oscillator output level vs. filament battery potential.

range of frequencies the change introduced in this manner is accurately proportional to the added reactance. Series inductance, of course, lowers the frequency, while series capacitance raises it. The stability requirement imposed on the adjusting reactance is only moderate, for its total effect upon the frequency should not be more than a few parts in a million.

The frequency measurements here presented were obtained using apparatus similar in principle to the frequency comparison equipment of the National Bureau of Standards.[5] Frequency differences

[5] "Harmonic Method of Intercomparing the Oscillators of the National Standard of Radio Frequency," E. G. Lapham, *Journal of Research of the National Bureau of Standards*, October 1936, p. 491.

between the oscillator under test and a reference bridge oscillator were read upon a linear scale calibrated directly in terms of frequency deviation. Full scale could be made one part in 10^4, 10^5, 10^6 or 10^7 by means of a simple switching operation. For most of the measurements in this paper the full-scale reading was one part in a million, and the resolution about $\pm$ 0.005 part in a million.

By using a recording meter with this measuring set, continuous frequency comparisons between two independent bridge oscillators

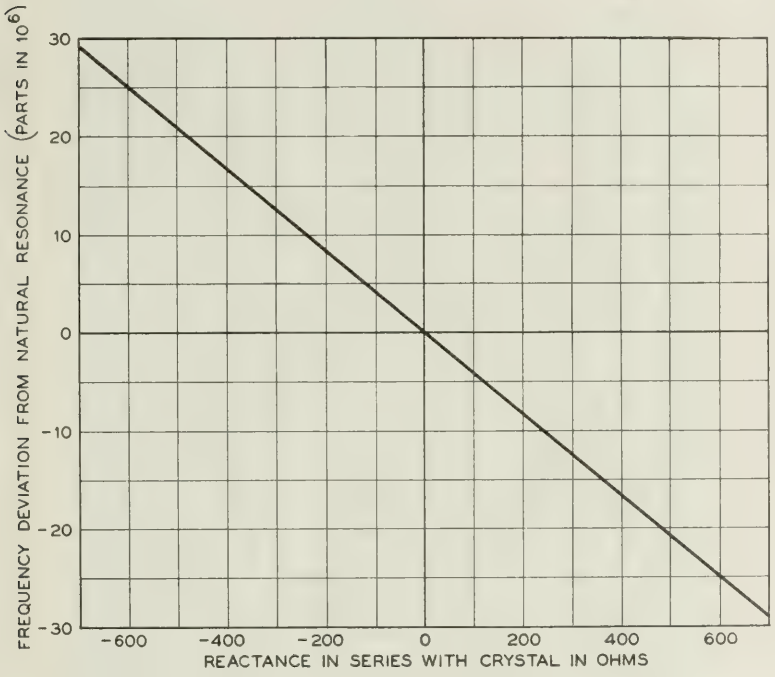


Fig. 10—Frequency of oscillator vs. adjusting reactance.

were obtained over a period of several months. Figure 11 is a photograph of a section of this record. It shows the short-time variations of both oscillators plus a small amount of scattering caused by the measuring equipment itself. The crystals were temperature-controlled in separate ovens, and the power was supplied from separate sets of laboratory batteries controlled to about $\pm$ 2 per cent in voltage. Shielding was ample to avoid any tendency to lock in step.

In addition to these small short-time variations, the oscillators exhibited a very slow upward drift in frequency, attributed to aging

of the mounted crystals. This aging decreased in a regular manner with time, the mean drift of one of the crystals being less than one part in ten million per month after three months of continuous operation,

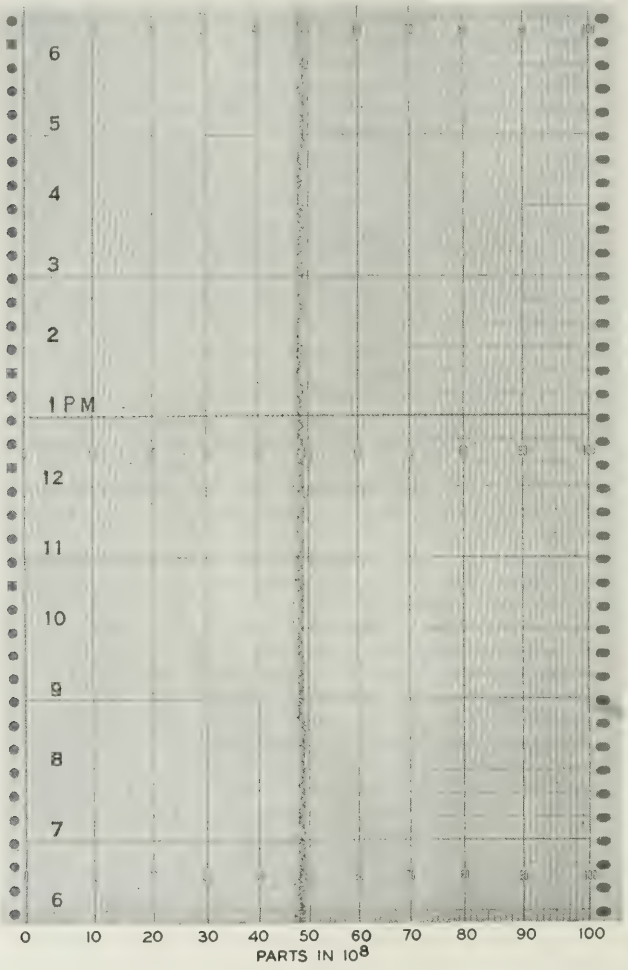


Fig. 11—Record of frequency comparison between two bridge stabilized oscillators. Full scale one part in a million. Variations less than ± 2 parts in one hundred million.

and about a third of this amount after seven months. In most applications, gradual frequency drift is not objectionable even though the required accuracy is very high, for readjustment is merely a matter of setting a calibrated dial.

Application

The bridge stabilized oscillator promises to become a useful tool in many commercial fields as well as in certain purely scientific problems such as time determination and physical and astronomical measurement. It may be used either to increase the frequency precision in applications where operating conditions are accurately controlled, or else to make such control unnecessary, affording high stability in spite of unfavorable conditions.

An interesting application in the field of geophysics has already been made in the form of a "Crystal Chronometer." This chronometer consists of a single-tube bridge oscillator, a frequency dividing circuit, and a synchronous timing motor. It was recently loaned by Bell Telephone Laboratories to the American Geophysical Union and was used with the Meinesz gravity-measuring equipment on a submarine gravity-survey expedition in the West Indies. Although operating under somewhat adverse conditions of power supply, temperature, and vibration, it was reported [6] to be more stable than any timing device previously available, errors in the gravity measurements introduced by the chronometer being negligibly small.

[6] "Gravity Measurement on the U. S. S. Barracuda," M. Ewing, and "Crystal Chronometer Time in Gravity Surveys," A. J. Hoskinson; pp. 66 and 77 resp., *Transactions of the American Geophysical Union*, Part I, 1937.

Effect of Space Charge and Transit Time on the Shot Noise in Diodes

By A. J. RACK

The theoretical analysis of the effect of space charge upon the "shot noise" in a planar diode shows that for practically all operating conditions, the tube noise is equivalent to the thermal resistance noise of the plate resistance at 0.644 times the cathode temperature. Noise in diodes of other than planar shapes is discussed and it is concluded that the same relation holds. It is shown that transit time produces the same high frequency modification for both the thermal and shot tube noise, and that the tube noise is decreased by transit time.

IN the study of noise in vacuum tubes, the effect of the space charge upon the shot noise has been a subject of considerable interest and practical importance. Several papers have been written in which it is shown that the shot noise is decreased by the space charge, and that the tube noise in a diode with space charge is equivalent to the thermal resistance noise of the plate impedance at a temperature slightly greater than half of that of the cathode.[1, 2, 3, 4] The most comprehensive analysis was made by Schottky and Spenke. These authors, employing a different method from the one here presented, have obtained the same general conclusions given in this paper, although they prefer to express the result in the form of a modified shot-noise equation, whereas for reasons developed below, the writer prefers the thermal form. The theoretical analysis and discussion presented here was undertaken to show in more detail the extent of the range of the operating condition for which the thermal resistance equivalent of tube noise is valid and to study the effect of transit time upon both the shot and thermal tube noise.

For convenience, the paper is divided into three parts. In the first section is given an exact mathematical treatment of the tube noise at low frequencies in a parallel plane diode for any degree of space charge. A discussion of the final tube noise equation obtained through this analysis, and the extension of these results for the planar diode to any other shape diode is given in Part II, where the presentation is such that the section may be read independently of the theoretical analysis in Part I. Through several approximations, Part III treats the effect of transit time upon tube noise in the planar diode.

Part I—General Low Frequency Analysis

In the development of the general equations for the direct current in vacuum tubes with space charge, account has been taken of the fact that the electrons are emitted from the cathode with Maxwellian velocity distribution. This fact has been verified experimentally by Germer,[5] and the resulting equations for the relation between current and voltage have been derived and investigated by Fry,[6] Langmuir,[7] and others. In the extension of this analysis to tube noise, it is only necessary to assume that the number of electrons emitted with any velocity does not remain constant, but fluctuates with time according to the well-known laws of probability. In the analysis on this basis, the frequencies involved will be considered to be sufficiently low so that any transit time effect is negligible.

Below is given a list of the definitions of various symbols to be used in the tube noise study of a parallel plane diode. The practical system of units is employed throughout.

$n(u_c)du_c$ = instantaneous rate of emission per unit area of the cathode of electrons with initial velocities between u_c and $u_c + du_c$ in the x-direction, regardless of the velocity components in the other directions,

$= n_0(u_c)du_c + \delta(u_c)du_c$,

$n_0(u_c)du_c$ = average rate of emission of electrons with x-directed velocities between u_c and $u_c + du_c$,

$\delta(u_c)du_c$ = instantaneous deviation from average rate of emission,

I = instantaneous anode current per unit area,

V = instantaneous potential with respect to cathode of a plane at a distance x from the cathode,

V' = instantaneous potential with respect to cathode of the potential minimum,

u = instantaneous velocity at x-plane of electrons which had an initial x-directed velocity of u_c at the cathode,

x' = instantaneous position of potential minimum,

e = charge on electron $= -1.59 \times 10^{-19}$ coulombs,

m = mass of electron $= 9.01 \times 10^{-28}$ grams,

h = ratio of dyne cms. to joules $= 10^{-7}$,

ϵ = permittivity of a vacuum in practical units $= 8.85 \times 10^{-14}$ farads/cm.,

k = Boltzmann's gas constant $= 1.372 \times 10^{-23}$ watts/degree Kelvin,

N = average total number of electrons emitted per second per unit area from the cathode,

T = absolute temperature of the cathode.

In the following analysis, it is assumed that the electrodes of the planar diode are infinite in extent, and that the electron emission is random, so that the equipotential surfaces are parallel planes perpendicular to the x-axis.

The potential distribution in such a planar diode operating with space charge is shown in Fig. 1. The origin of coordinates is taken at the cathode, and the potential minimum formed by space charge occurs at a distance $x = x'$ from the cathode. The subscript α will be used to denote the space between cathode and potential minimum while β applies similarly to the space between minimum and anode. Of all

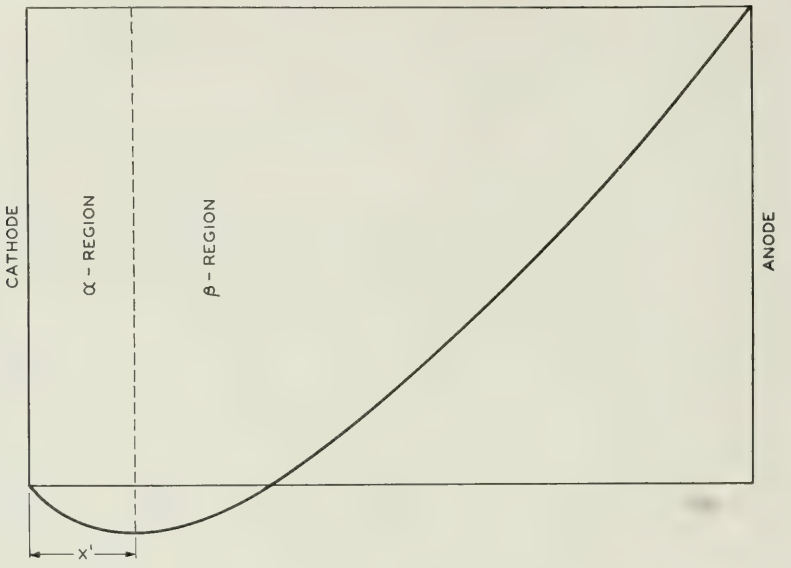


Fig. 1—Potential distribution in planar diode.

the electrons emitted from the cathode only those whose x-velocity exceeds the value u_c' corresponding to the potential minimum can penetrate the barrier and proceed to the anode. Electrons with lesser values of initial velocity will come to rest at a point in the α-region where the potential corresponds to their initial velocity and will then return to the cathode. The anode current density is thus given by

$$I = e \int_{u_c'}^{\infty} n(u_c) du_c, \tag{1}$$

while the relation between velocity u and potential V at a given value of x is

$$u^2 = u_c^2 - \frac{2e}{hm} V. \tag{2}$$

A third fundamental relation is Poisson's equation which becomes in the parallel plane case under consideration

$$\epsilon \frac{d^2V}{dx^2} = -\rho. \tag{3}$$

In the α-region the total charge density is made up of three classes of electrons, namely

1. Those destined to pass the potential minimum and arrive at the anode.
2. Those moving away from the cathode but which will not travel as far as the minimum point.
3. Those returning to the cathode.

Corresponding to each class of electrons, there is an associated current, ρu, so that each of the three densities ρ_1, ρ_2 or ρ_3, may be expressed by a relation of the form,

$$\rho_n = \frac{I_n}{u}. \tag{4}$$

When it is remembered that the potential and velocity at a given value of x are uniquely related through (2), then it is easy to see that the total density for a given plane in the α-region is given by

$$\rho_\alpha = e\int_{u_c'}^{\infty} \frac{n(u_c)}{u}\,du_c + 2e\int_{v}^{u_c'} \frac{n(u_c)}{u}\,du_c, \tag{5}$$

where the first term represents the contribution of electrons in class 1 above, while the second term represents the contribution of electrons in classes 2 and 3. The contribution of class 3 is equal to that of class 2. The lower integration limit v of the second term of (5) represents the initial velocity of an electron which would just arrive at the value of x under consideration before coming to rest and starting back toward the cathode and the limit u_c' in both terms represents the initial velocity of an electron which comes to rest just at the potential minimum. Thus, from (2)

$$v = \sqrt{\frac{2e}{hm}\,V} \qquad \text{and} \qquad u_c' = \sqrt{\frac{2e}{hm}\,V'}. \tag{6}$$

In the β-region there is only one class of electrons, so that the density is more simply expressed. Thus,

$$\rho_\beta = e\int_{u_c'}^{\infty} \frac{n(u_c)}{u}\,du_c. \tag{7}$$

The value of ρ in (5) and (7) may each be expressed in terms of d^2V/dx^2 by the use of (3), and the integration of these two Poisson's relations for the common boundary condition that the electric force is

zero at the potential minimum has the following result:

$$\frac{(dV_\alpha)^2}{(dx)} = \frac{2hm}{\epsilon}\int_{u_{c'}}^{\infty}(u - u')n(u_c)du_c + \frac{4hm}{\epsilon}\int_{v}^{u_{c'}} un(u_c)du_c, \tag{8}$$

$$\frac{(dV_\beta)^2}{(dx)} = \frac{2hm}{\epsilon}\int_{u_{c'}}^{\infty}(u - u')n(u_c)du_c, \tag{9}$$

where u' is the electronic velocity at the potential minimum, i.e., $(u')^2 = u_c^2 - (2e/hm)V'$.

At this point the analysis departs for the first time from the classic analyses of Fry [6] and Langmuir,[7] through the introduction of the concept that the instantaneous rate of emission may be expressed as the sum of an average rate of emission plus an instantaneous deviation. That is,

$$n(u_c) = n_0(u_c) + \delta(u_c), \tag{10}$$

transforms (8) and (9) into the following equations:

$$\frac{(dV_\alpha)^2}{(dx)} = \frac{2hm}{\epsilon}\int_{u_{c'}}^{\infty}(u - u')n_0(u_c)du_c + \frac{4hm}{\epsilon}\int_{v}^{u_{c'}} un_0(u_c)du_c + \alpha(\delta), \tag{11}$$

where

$$\alpha(\delta) = \frac{2hm}{\epsilon}\int_{u_{c'}}^{\infty}(u - u')\delta(u_c)du_c + \frac{4hm}{\epsilon}\int_{v}^{u_{c'}} u\delta(u_c)du_c$$

and

$$\frac{(dV_\beta)^2}{(dx)} = \frac{2hm}{\epsilon}\int_{u_{c'}}^{\infty}(u - u')n_0(u_c)du_c + \beta(\delta), \tag{12}$$

where

$$\beta(\delta) = \frac{2hm}{\epsilon}\int_{u_{c'}}^{\infty}(u - u')\delta(u_c)du_c.$$

Since the average rate of emission may be expressed by the Maxwellian relation,

$$n_0(u_c) = 2\alpha N u_c \epsilon^{-\alpha u_c^2},$$

where

$$\alpha = \frac{hm}{2kT},$$

the indicated integrations in (11) and (12) have as a result,

$$\frac{(kT)^2}{(e)}\frac{(d\eta_\alpha)^2}{(dx)} = \frac{Nhm}{\epsilon}\sqrt{\frac{\pi}{\alpha}}\epsilon^{-\eta'} \times \left[\epsilon^{\eta} - 1 + \epsilon^{\eta}P(\sqrt{\eta}) - 2\sqrt{\frac{\eta}{\pi}}\right] + \alpha(\delta) \tag{13}$$

and

$$\frac{(kT)^2}{(e)}\frac{(d\eta_\beta)^2}{(dx)} = \frac{Nhm}{\epsilon}\sqrt{\frac{\pi}{\alpha}}\epsilon^{-\eta'} \times \left[\epsilon^{\eta} - 1 - \epsilon^{\eta}P(\sqrt{\eta}) + 2\sqrt{\frac{\eta}{\pi}}\right] + \beta(\delta), \quad (14)$$

where

$$\eta = \frac{e}{kT}(V' - V), \qquad \eta' = \frac{eV'}{kT},$$

$$P(x) = \frac{2}{\sqrt{\pi}}\int_0^x \epsilon^{-x^2}dx.$$

The fact that both $\alpha(\delta)$ and $\beta(\delta)$ are very small greatly simplifies the solution for the distance coordinate x in (13) and (14). The process is to invert the two equations, respectively, extract the square root, and then expand the right-hand side in powers of $\alpha(\delta)$ and $\beta(\delta)$, respectively. This results in expressions for $dx/d\eta$ which can be integrated term by term. However, the small values of $\alpha(\delta)$ and $\beta(\delta)$ allow powers higher than the first to be disregarded, and hence,

$$\frac{e}{kT}x' = \frac{F(\eta')}{\left[\frac{Nhm}{\epsilon}\sqrt{\frac{\pi}{\alpha}}\epsilon^{-\eta'}\right]^{1/2}} - \frac{1}{2\left[\frac{Nhm}{\epsilon}\sqrt{\frac{\pi}{\alpha}}\epsilon^{-\eta'}\right]^{3/2}}\int_0^{\eta'}\frac{\alpha(\delta)d\eta}{\Phi(\eta)} \quad (15)$$

and

$$\frac{e}{kT}(x - x') = \frac{f(\eta)}{\left[\frac{Nhm}{\epsilon}\sqrt{\frac{\pi}{\alpha}}\epsilon^{-\eta'}\right]^{1/2}} - \frac{1}{2\left[\frac{Nhm}{\epsilon}\sqrt{\frac{\pi}{\alpha}}\epsilon^{-\eta'}\right]^{3/2}}\int_0^{\eta}\frac{\beta(\delta)d\eta}{\varphi(\eta)}, \quad (16)$$

where for convenience

$$\left.\begin{aligned} F(\eta') &= \int_0^{\eta'}\frac{d\eta}{\left[\epsilon^{\eta} - 1 + \epsilon^{\eta}P(\sqrt{\eta}) - 2\sqrt{\frac{\eta}{\pi}}\right]^{1/2}} \\ f(\eta) &= \int_0^{\eta}\frac{d\eta}{\left[\epsilon^{\eta} - 1 - \epsilon^{\eta}P(\sqrt{\eta}) + 2\sqrt{\frac{\eta}{\pi}}\right]^{1/2}} \\ \Phi(\eta) &= \left[\epsilon^{\eta} - 1 + \epsilon^{\eta}P(\sqrt{\eta}) - 2\sqrt{\frac{\eta}{\pi}}\right]^{3/2} \\ \varphi(\eta) &= \left[\epsilon^{\eta} - 1 - \epsilon^{\eta}P(\sqrt{\eta}) + 2\sqrt{\frac{\eta}{\pi}}\right]^{3/2} \end{aligned}\right\}. \quad (17)$$

Up to this point, the present analysis is very similar to that given by Spenke.[3] For the reasons stated above, the two methods digress hereafter.

Since the instantaneous position of the potential minimum depends upon the operating conditions as indicated above, the elimination of this variable by the addition of the two equations (15) and (16) results in a simpler expression for the cathode-anode spacing, namely:

$$\frac{e}{kT}x\left[\frac{Nhm}{\epsilon}\sqrt{\frac{\pi}{\alpha}}\epsilon^{-\eta'}\right]^{1/2} = F(\eta') + f(\eta) - \frac{1}{2\left[\frac{Nhm}{\epsilon}\sqrt{\frac{\pi}{\alpha}}\epsilon^{-\eta'}\right]}\left\{\int_0^{\eta'}\frac{\alpha(\delta)}{\Phi(\eta)}d\eta + \int_0^{\eta}\frac{\beta(\delta)}{\varphi(\eta)}d\eta\right\}. \quad (18)$$

To separate the noise or fluctuation component of the potentials from their average values, it is necessary to assume at any plane in the diode that it is possible to express both the instantaneous voltage and velocity as a steady state or average value plus a very small superimposed fluctuation component. Since this assumption does not result in any discontinuities, it is possible to express the instantaneous values of the dimensionless variable as

$$\left.\begin{aligned}\eta &= \eta_0 + \eta_1\\ \text{and}\qquad \eta' &= \eta_0' + \eta_1',\end{aligned}\right\} \quad (19)$$

where both η_1 and η_1' are very small. From this assumption, it may readily be shown that the d-c. solution and the first approximations to the fluctuation component of the solution for (18) are respectively

$$\frac{ex}{kT}\left[\frac{Nhm}{\epsilon}\sqrt{\frac{\pi}{\alpha}}\epsilon^{-\eta_0'}\right]^{1/2} = F(\eta_0') + f(\eta_0) \quad (20)$$

and

$$-\frac{ex}{kT}\left[\frac{Nhm}{\epsilon}\sqrt{\frac{\pi}{\alpha}}\epsilon^{-\eta_0'}\right]^{1/2}\frac{\eta_1'}{2} = \eta_1'\frac{dF(\eta_0')}{d\eta_0'} + \eta_1\frac{df(\eta_0)}{d\eta_0} - \frac{1}{2\left[\frac{Nhm}{\epsilon}\sqrt{\frac{\pi}{\alpha}}\epsilon^{-\eta_0'}\right]}\left[\int_0^{\eta_0'}\frac{\alpha(\delta)}{\Phi(\eta)}d\eta + \int_0^{\eta_0}\frac{\beta(\delta)}{\varphi(\eta)}d\eta\right]. \quad (21)$$

In the last equation, the average or d-c. values of all quantities except η_1, η_1' and $\delta(u_c)$ are to be used.

To avoid the necessity of using long, awkward equations, it will be of great convenience to define several new variables.

Let

$$\left.\begin{aligned} y &= \sqrt{\alpha}\, u' \\ B &= \frac{F(\eta_0') + f(\eta_0)}{2} + \frac{dF(\eta_0')}{d\eta_0'} + \frac{df(\eta_0)}{d\eta_0} \\ C &= \frac{1}{\sqrt{\pi}} \int_0^{\eta_0'} \frac{\left[\sqrt{y^2 + \eta} - y\right]}{\Phi(\eta)}\, d\eta \\ D &= \frac{1}{\sqrt{\pi}} \int_0^{\eta_0} \frac{\left[\sqrt{y^2 + \eta} - y\right]}{\varphi(\eta)}\, d\eta \end{aligned}\right\}. \tag{22}$$

From the definition for η, given in connection with (13) and (14), it may be shown that

$$\eta_1 = \eta_1' - \frac{e}{kT} V_1, \tag{23}$$

where V_1 is the a-c. anode potential. With the definition given for $\alpha(\delta)$ and $\beta(\delta)$ in (11) and (12), and from the above relations, (21) may be rewritten as

$$\eta_1' B - \frac{eV_1}{kT}\frac{df(\eta_0)}{d\eta_0} = \frac{1}{N\epsilon^{-\eta_0'}} \left\{ \int_{\bar{u}_c}^{\infty} (C + D)\delta(u_c)du_c + \frac{2}{\sqrt{\pi}} \int_0^{\eta_0'} \int_{\bar{\tau}}^{\bar{u}_c} \frac{\sqrt{y^2 + \eta}\, \delta(u_c)du_c d\eta}{\Phi(\eta)} \right\}, \tag{24}$$

where

$$\bar{u}_c = \sqrt{\frac{2e}{hm} V_0'}, \qquad \bar{\tau} = \sqrt{\frac{2e}{hm} V_0}.$$

Before (24) can be used, the relation between η_1' and the a-c. anode current, and also the expression for the a-c. plate impedance must be obtained. From the general relation given in (1), the instantaneous current per unit area is

$$I = e\int_{u_c'}^{\infty} n(u_c)du_c = e\int_{u_c'}^{\infty} n_0(u_c)du_c + e\int_{u_c'}^{\infty} \delta(u_c)du_c$$
$$= Ne\epsilon^{-\eta'} + e\int_{u_c'}^{\infty} \delta(u_c)du_c.$$

Since the instantaneous voltage of the potential minimum may be expressed as a sum of a d-c. component plus a small a-c. fluctuation, the d-c. and first order fluctuation plate current are respectively

$$I_0 = Ne\epsilon^{-\eta_0'}, \tag{25}$$

$$I_1 = -\eta_1' Ne\epsilon^{-\eta_0'} + e\int_{\bar{u}}^{\infty} \delta(u_c)du_c. \tag{26}$$

Whence,

$$\eta_1' = \frac{e}{I_0}\int_{u_c}^{\infty} \delta(u_c)du_c - \frac{I_1}{I_0}. \tag{27}$$

This is the desired relation between η_1' and the a-c. anode current.

To find the other desired relation, it is first observed that by (20) and (25), the d-c. voltage-current relation is

$$\frac{ex}{kT}\left[\frac{hm}{\epsilon e}\sqrt{\frac{\pi}{\alpha}}\right]^{1/2} I_0^{1/2} = F(\eta_0') + f(\eta_0). \tag{28}$$

This is identical with that obtained by Fry and Langmuir. The values of $F(\eta_0')$ and $f(\eta_0)$ have been tabulated by Langmuir.[7]

From this d-c. relation, it may readily be shown that the a-c. plate impedance of the planar diode is

$$r_p = \frac{\frac{F(\eta_0') + f(\eta_0)}{2} + \frac{dF(\eta_0')}{d\eta_0'} + \frac{df(\eta_0)}{d\eta_0}}{\frac{eI_0}{kT}\frac{df(\eta_0)}{d\eta_0}} = \frac{BkT}{eI_0}\frac{1}{\frac{df(\eta_0)}{d\eta_0}}. \tag{29}$$

Since the noise generator voltage,

$$E = I_1 r_p + V_1,$$

the substitution of (27) and (29) reduces (24) to the following relations:

$$\begin{aligned}\frac{e}{kT}(I_1 r_p + V_1)\frac{df(\eta_0)}{d\eta_0} &= \frac{eE}{kT}\frac{df(\eta_0)}{d\eta_0} = \frac{e}{I_0}\left\{\int_{u_c}^{\infty}(B - C - D)\delta(u_c)du_c\right. \\ &\left. - \frac{2}{\sqrt{\pi}}\int_0^{\eta_0'}\int_{-}^{u_c}\frac{\sqrt{y^2+\eta}\,\delta(u_c)du_c d\eta}{\Phi(\eta)}\right\}. \end{aligned} \tag{30}$$

By the additional symbols,

$$\left.\begin{aligned} H(u_c, \eta_0, \eta_0') &\equiv \frac{kT}{I_0\frac{df(\eta_0)}{d\eta_0}}(B - C - D) \\ G(u_c, \eta, \eta_0) &\equiv -\frac{2}{\sqrt{\pi}}\frac{kT}{I_0\frac{df(\eta_0)}{d\eta_0}}\frac{\sqrt{y^2+\eta}}{\Phi(\eta)} \end{aligned}\right\}, \tag{31}$$

the noise generator voltage may be expressed in the condensed form,

$$E = \int_{\ddot{u}_c}^{\infty} H\delta(u_c)du_c + \int_0^{\eta_0'}\int_{\bar{v}}^{\ddot{u}_c} G\delta(u_c)du_c d\eta. \tag{32}$$

Unfortunately (32) cannot be integrated because the specific value of the instantaneous deviation in the rate of electron emission is unknown. Moreover, as shown by Fry [10] there exists no frequency spectrum for this deviation. The reason is because there is no way of foretelling at a particular instant just when the next electron is going to be emitted from a thermionic cathode. It does not follow, however, that Fourier methods are powerless, as the following argument will show. Imagine that the emission in a thermionic system has been going on for a long time, and place a recorder in the system which makes an oscillograph record of the voltage produced across the tube by the fluctuating current. Let the record be made over a long period of time. Then, it is a perfectly possible thing to analyze the record so obtained into a Fourier spectrum. The result will give no information that the Fourier spectrum which would be obtained on the oscillograph during an ensuing time period of equal length would be the same as the one which has been secured during the past. However, when the mean square value of the Fourier spectrum for the recorded interval is computed, it is found that the mean square value of any two records so obtained is the same when they are both produced by random emission. Moreover, the mean square value within a specified frequency interval is also the same in the two records. These facts result from the random character of the events producing the records, as may be seen even more clearly by examination of the mathematical steps in the following equations, particularly as given in the progression from (37) to (38). It follows, then, that one is justified in concluding that the mean square response of an electrical system to a random excitation can be calculated by the Fourier series method, and that the result so obtained applies equally well either to systems which have been measured in the past, or to those which will be measured in the future, provided only that they both are similar in their configuration and external operating conditions.

To obtain the Fourier expression for the emission deviation, it will be assumed that this function repeats itself after a very long period of time. To find this Fourier Series for the instantaneous deviation from the average rate of emission of electrons with x-directed velocities between u_c and $u_c + du_c$, the very long period of time, L, is divided into P equal intervals of length, τ, where τ is assumed to be mathematically small. The exact number of electrons emitted during any of these intervals of time with velocities between u_c and $u_c + du_c$ per unit area of the cathode is denoted by $n_m(u_c)\tau$. The average rate of

emission of electrons in this velocity class may then be defined as follows:

$$n_0(u_c) \equiv \frac{1}{P} \sum_{m=1}^{P} n_m(u_c). \tag{33}$$

During each interval, the deviation from the average rate of emission is

$$\delta_m(u_c) = n_m(u_c) - n_0(u_c). \tag{34}$$

The mean square of the instantaneous deviation for all P intervals is then

$$\overline{\delta^2(u_c)} = \frac{1}{P} \sum_{m=1}^{P} [n_m(u_c) - n_0(u_c)]^2. \tag{35}$$

The value of the mean square emission deviation may be found from the following considerations. In previous studies of pure shot noise, it was assumed that the electrons are emitted from the cathode independently of one another, that is, the probability of an electron being emitted during a very small interval of time depends only upon the average rate of emission and the length of the time interval. The classical theoretical equation developed for the shot noise from this assumption was found experimentally to be correct to well within experimental errors. Hence, the same assumption may be made in this analysis of tube noise as the presence of the space charge near the cathode can have but a negligible effect upon the rate of emission of electrons from the cathode. Thus, since the electrons in any velocity class are also emitted independently of one another, the application of probability theory shows that the mean square deviation in the total number of electrons emitted with velocities between u_c and $u_c + du_c$ in a time, τ, is equal to the average number of electrons of this velocity class emitted in the same time, τ.

That is,

$$\frac{1}{P} \sum_{m=1}^{P} \tau^2 \delta_m{}^2(u_c) = \tau n_0(u_c),$$

or

$$\sum_{m=1}^{P} \delta_m{}^2(u_c) = \frac{L}{\tau^2} n_0(u_c). \tag{36}$$

Since the period of time, τ, is mathematically small, it can be assumed that the instantaneous deviation from the average rate of emission is constant in each of the P intervals and equal to $\delta_m(u_c)$. If it is assumed that the function representing the instantaneous deviation repeats itself after a very long period of time, L, the Fourier series for the P square-top pulses comprising the instantaneous

deviation function may be shown to be:

$$\delta(u_c) = \sum_{l=-\infty}^{\infty} \sum_{m=1}^{P} \frac{\tau \delta_m(u_c)}{L} \left[\frac{e^{-il\omega\tau} - 1}{-il\omega\tau} \right] e^{il\omega(t-t_m)}, \tag{37}$$

where $\omega = 2\pi/L$, and $t_m = m\tau$.

The mean square deviation may be found by squaring the above expression and averaging the result over the long period of time L. The result is

$$\overline{\delta^2(u_c)} = 2 \sum_{l=1}^{\infty} \sum_{k=1}^{P} \sum_{m=1}^{P} \frac{\tau^2}{L^2} \delta_m(u_c) \delta_k(u_c) \left[\frac{e^{-il\omega\tau} - 1}{-il\omega\tau} \right] \times \left[\frac{e^{+il\omega\tau} - 1}{+il\omega\tau} \right] e^{-il\omega(t_k - t_m)}.$$

Since the time-average of the instantaneous emission deviation must be zero over the period, L, the contribution to the mean square from the double summation with respect to k and m is zero, unless $m = k$.

Thus

$$\overline{\delta^2(u_c)} = 4 \sum_{l=1}^{\infty} \frac{\tau^2}{L^2} \left[\frac{1 - \cos l\omega\tau}{l^2\omega^2\tau^2} \right] \sum_{m=1}^{P} \delta_m{}^2(u_c). \tag{38}$$

From (36), this equation reduces to

$$\overline{\delta^2(u_c)} = \frac{4}{L} \sum_{l=1}^{\infty} \left[\frac{1 - \cos l\omega\tau}{l^2\omega^2\tau^2} \right] n_0(u_c). \tag{39}$$

The contribution to the mean square of the instantaneous deviation in the electron emission, from the frequencies between f and $f + df$ is given by

$$\overline{\delta_f{}^2(u_c)} = \frac{4}{L} \sum_{l=2\pi f/\omega}^{l=2\pi(f+df)/\omega} \left[\frac{1 - \cos l\omega\tau}{l^2\omega^2\tau^2} \right] n_0(u_c). \tag{40}$$

The limit of this expression as the length of the periodicity, L, is made infinite, may readily be shown to be

$$\overline{\delta_f{}^2(u_c)} = 2n_0(u_c)df. \tag{41}$$

It is now possible to proceed to find the mean square value of the noise generator voltage given in (32) with the aid of (37) and (41). From (37), the noise generator voltage may be expressed as

$$E = \sum_{l=-\infty}^{\infty} \sum_{m=1}^{P} \frac{\tau}{L} \left[\frac{e^{-il\omega\tau} - 1}{-il\omega\tau} \right] e^{il\omega(t-t_m)} \times \left\{ \int_{\bar{u}}^{\infty} H(u_c) \delta_m(u_c) du_c + \int_0^{\eta_0'} \int_{\sqrt{\alpha(\eta-\eta_0')}}^{\bar{u}_c} G \delta_m(u_c) du_c d\eta \right\}. \tag{42}$$

The mean square of this equation for the noise generator voltage may be obtained by finding the average of the square of the expression over a very long period of time; that is

$$\overline{E^2} = 4\sum_{l=1}^{\infty}\sum_{k=1}^{P}\sum_{m=1}^{P}\frac{\tau^2}{L^2}\left[\frac{1-\cos l\omega\tau}{l^2\omega^2\tau^2}\right]e^{il\omega(t_m-t_k)}$$
$$\times\left\{\int_{x=\bar{u}_c}^{\infty}\int_{y=\bar{u}_c}^{\infty}H(x)H(y)\delta_m(x)\delta_k(y)dxdy\right.$$
$$+2\int_{\eta=0}^{\eta_0'}\int_{y=\bar{u}_c}^{\infty}\int_{x=\sqrt{\alpha(\eta-\eta_0')}}^{\infty}G(\eta,x)H(y)\delta_m(x)\delta_k(y)dxdyd\eta$$
$$+\int_{\eta=0}^{\eta_0'}\int_{z=0}^{\eta_0'}\int_{y=\sqrt{\alpha(z-\eta_0')}}^{\bar{u}_c}\int_{x=\sqrt{\alpha(\eta-\eta_0')}}^{\bar{u}_c}G(\eta,x)G(z,y)$$
$$\left.\times\,\delta_m(x)\delta_k(y)dxdydzd\eta\right\}. \quad (43)$$

In the above equation, the contribution to the mean square noise generator voltage from the summation with respect to m, for a fixed value of k, is zero unless $m = k$, since the long time average of the emission deviation must be zero. Furthermore, since the electrons are emitted independently of one another

$$\sum_{m=1}^{P}\delta_m(x)\delta_m(y) = 0, \quad \text{unless} \quad x = y.$$

From these considerations, the contribution to the mean square generator voltage from the second integral in (43) is zero since x and y have no common value. The contribution from the last integral is a bit more difficult to obtain. However, from (41), the contribution to the mean square noise generator voltage from the frequencies between f and $f + df$ can be shown to be

$$\overline{E_f^2} = 2df\left\{\int_{u_c}^{\infty}H^2(u_c)n_0(u_c)du_c\right.$$
$$+\int_{\eta=0}^{\eta_0'}\int_{z=0}^{\eta}\int_{u_c=\sqrt{\alpha(\eta-\eta_0')}}^{\bar{u}_c}G(\eta,u_c)G(z,u_c)n_0(u_c)du_cdzd\eta$$
$$\left.+\int_{\eta=0}^{\eta_0'}\int_{z=\eta}^{\eta_0'}\int_{u_c=\sqrt{\alpha(z-\eta_0')}}^{\bar{u}_c}G(\eta,u_c)G(z,u_c)n_0(u_c)du_cdzd\eta\right\}. \quad (44)$$

In terms of the variable $y = \sqrt{\alpha}u'$, the average rate of emission may readily be shown to be

$$n_0(u_c)du_c = \frac{2I_0}{e}ye^{-y^2}dy.$$

From this value of the average rate of emission, and from the definition of H and G in (31), and since the plate impedance of the planar diode is given by (28), the final expression for the mean square noise generator voltage may be expressed as follows:

$$\overline{E_f^2} = 4kr_p(\lambda T)df, \tag{45}$$

where

$$\begin{aligned}\lambda = \frac{1}{B\dfrac{df(\eta_0)}{d\eta_0}}\Bigg\{ &\int_0^\infty y(B - C - D)^2 e^{-y^2}dy \\ &+ \frac{4}{\pi}\int_{z=0}^{\eta_0'}\int_{x=0}^{z}\int_{y=\sqrt{-x}}^{0} \frac{y\sqrt{y^2+x}\,\sqrt{y^2+z}\,e^{-y^2}}{\Phi(x)\Phi(z)}\,dxdydz \\ &+ \frac{4}{\pi}\int_{z=0}^{\eta_0'}\int_{x=z}^{\eta_0'}\int_{y=\sqrt{-z}}^{0} \frac{y\sqrt{y^2+x}\,\sqrt{y^2+z}\,\epsilon^{-y^2}}{\Phi(x)\Phi(z)}\,dxdydz\Bigg\},\end{aligned} \tag{46}$$

$$\left.\begin{aligned}
&\eta_0 = \frac{e}{kT}(V' - V_p), \qquad \eta_0' = \frac{e}{kT}V' \\
&B = \frac{F(\eta_0') + f(\eta_0)}{2} + \frac{dF(\eta_0')}{d\eta_0'} + \frac{df(\eta_0)}{d\eta_0}, \qquad r_p = \frac{B}{\dfrac{eI_0}{kT}\dfrac{df(\eta_0)}{d\eta_0}} \\
&F(\eta_0') = \int_0^{\eta_0'} \frac{dx}{\left[\epsilon^x - 1 + \epsilon^x P(\sqrt{x}) - 2\sqrt{\frac{x}{\pi}}\right]^{1/2}}, \\
&f(\eta_0) = \int_0^{\eta_0} \frac{dx}{\left[\epsilon^x - 1 - \epsilon^x P(\sqrt{x}) + 2\sqrt{\frac{x}{\pi}}\right]^{1/2}} \\
&C = \frac{1}{\sqrt{\pi}}\int_0^{\eta_0'} \frac{[\sqrt{y^2+x} - y]}{\Phi(x)}\,dx, \\
&D = \frac{1}{\sqrt{\pi}}\int_0^{\eta_0} \frac{[\sqrt{y^2+x} - y]dx}{\left[\epsilon^{-x} - 1 - \epsilon^x P(\sqrt{x}) + 2\sqrt{\frac{x}{\pi}}\right]^{3/2}} \\
&\Phi(x) = \left[\epsilon^x - 1 + \epsilon^x P(\sqrt{x}) - 2\sqrt{\frac{x}{\pi}}\right]^{3/2}, \\
&P(x) = \frac{2}{\sqrt{\pi}}\int_0^x \epsilon^{-x^2}dx.
\end{aligned}\right\} \tag{47}$$

Part II—General Discussion

The analysis in Part I shows that as soon as a potential minimum exists, the tube noise in a planar diode is equivalent to the thermal

noise of the plate resistance at an effective temperature which is a function of that of the cathode.

In general, the effective value of the diode plate resistance temperature for any operating condition is very difficult to obtain because of the complexity of the final noise equations (45) and (46). However, the limiting value of the ratio of the effective plate resistance temperature to that of the cathode, denoted by "λ" in (45), may be evaluated very readily for certain limiting conditions.

One encounters the first of these conditions when the plate potential and cathode emission are such that the potential minimum has moved just up to the cathode, and is in fact on the point of disappearing. This condition is secured by decreasing the space charge to values less than are required for the formation of a potential minimum away from the cathode. In the equations, it is represented by letting the quantity η_0' approach zero, where η_0' is the natural logarithm of the ratio of the saturation current to the anode current. For this set of operating conditions, all the electrons emitted from the cathode will go to the anode, and hence the condition is appropriate to the study of pure shot noise.

A second condition is obtained when the plate potential is equal in value to the potential of the minimum. Physically, this condition means that the minimum has moved just up to the anode, and requires a negative value for the plate potential. Mathematically, it is represented by a zero value for the quantity η_0, where η_0 is equal to the difference between η_0' and $(e/kT)V_p$. For negative plate voltages greater in magnitude than that of the potential minimum, all electrons having an initial kinetic energy greater than eV_p will reach the anode regardless of the presence of the space charge existing between the two electrodes. For these conditions, the diode becomes a temperature limited current device.

A third limiting condition occurs when the plate potential is large in magnitude compared with that of the potential minimum referred to the cathode. In this condition a potential minimum still exists. It is represented in the mathematics by letting the quantity η_0 become large. This condition represents the normal operating condition for the diode.

As the space charge is decreased, making η_0' very small, from (47), the diode plate impedance becomes very large through the action of $dF(\eta_0')/d\eta_0'$ which becomes infinite as η_0' approaches zero. As all other quantities involved remain finite, the mean square noise current for a very small space charge is

$$\overline{I_1^2} = \frac{\bar{E}_f^2}{(R_p + Z)^2} = \frac{4kTdf}{B\,\dfrac{kT}{eI_0}} \frac{\dfrac{df(\eta_0)}{d\eta_0}}{B\,\dfrac{df(\eta_0)}{d\eta_0}} B^2 \int_0^\infty y\epsilon^{-y^2}dy \to 2eI_0df. \quad (48)$$

Thus, as the potential minimum voltage is reduced to zero, the tube noise as given by (45) reduces to the well known shot effect equation.

For some space charge at the cathode, the value of λ in (45) has definite limiting values for both very low and for very large plate voltages. For a very small value of η_0, that is for negative plate voltage, the value of B defined in (47) is very large because $df(\eta_0)/d\eta_0$ becomes infinite as η_0 is decreased to zero. Thus as $\eta_0 \to 0$

$$\lambda \to \frac{1}{B^2} B^2 \int_0^\infty y\epsilon^{-y^2}dy = \frac{1}{2}. \quad (49)$$

Hence, for any value of space charge, the effective plate resistance temperature for negative plate voltages is one-half of the cathode temperature, under the restriction that no potential minimum exists between the cathode and anode.

Since the diode is usually operated with a positive plate voltage, the value of the effective plate resistance temperature for a large value of η_0 is of more interest. For η_0' not equal to zero, and a large value of plate voltage, it can be readily shown that the values of $f(\eta_0)$ and of D are much larger than any other quantities involved in the equation for λ. After a bit of mathematical operation, it may be shown that the limiting values for $f(\eta_0)$ and D are

$$D = \frac{\pi^{1/4}}{2^{3/2}}\left[\frac{4}{3}\eta_0^{3/4} + 3\sqrt{\pi}\eta_0^{1/4} + \cdots - (4y\eta_0^{1/4} + \cdots)\right]$$

$$f(\eta_0) = \frac{\pi^{1/4}}{\sqrt{2}}\left[\frac{4}{3}\eta_0^{3/4} + \sqrt{\pi}\eta_0^{1/4} + \cdots\right].$$

From these relations, the limiting value of λ for a large plate voltage is given by

$$\lambda = 3\int_0^\infty y\left[\sqrt{2}y - \sqrt{\frac{\pi}{4}}\right]^2 \epsilon^{-y^2}dy = 3\left(1 - \frac{\pi}{4}\right) = 0.644. \quad (50)$$

Thus, for any value of space charge, as long as a potential minimum exists, a sufficiently large value of plate voltage may always be found for which the effective plate resistance temperature is 0.644 times the cathode temperature.

It is possible to obtain a good approximation for the effective diode temperature for any operating condition by the following method. The values of "C" and "D" in (47) may be found without too much difficulty by graphical integration for several different values of y, η_0 and η_0'. From the tabulated values for $F(\eta_0')$ and $f(\eta_0)$ given by Langmuir, and from the values found for C and D, the integral,

$$S = \int_0^\infty y[B - C - D]^2 \epsilon^{-y^2} dy, \tag{51}$$

may be evaluated by mechanical means for several values of η_0 and η_0'. This gives the first integral in (46).

It was found practically impossible to calculate directly the contribution to λ from the last two integrals in (46). However, a rough approximation to them may be found indirectly by the following method: If the sum of the two integrals is denoted by Q, then (46) may be written

$$\lambda = \frac{S + Q}{B \dfrac{df(\eta_0)}{d\eta_0}},$$

or $Q = B\lambda(df(\eta_0)/d\eta_0) - S =$ function of η_0' only.

For a fixed value of η_0', the solution of the above equation for several values of η_0 should give a constant value for Q. Unfortunately, only the limiting values of λ are known. However, if the limiting value of 0.644 is substituted for λ in this equation, the calculated value of Q, for a fixed value of η_0', should approach a constant value as η_0 is increased since λ does assume the 0.644 value for η_0 sufficiently large. The limiting value of Q calculated in this manner is the desired contribution to λ from the last two integrals in (46). This method of evaluating Q cannot be very accurate since it involves the difference of two quantities of the same magnitude. However, since Q is small compared to the contribution from the first integral in (46), a large error in Q will introduce a much smaller error in the value of λ.

The values of the effective diode plate resistance temperature calculated in this manner for several different operating conditions are shown in Fig. 2. These curves indicate that the effective diode temperature is 0.644 times the cathode temperature for all practical operating conditions. The values of η_0' and η_0 may be determined from the following relations:

$$\eta_0' = \log_\epsilon \frac{I_s}{I_p}, \tag{52}$$

where I_s is the saturation current and I_p is the anode current, and

$$\eta_0 = \eta_0' - \frac{e}{kT} V_p = \eta_0' + \frac{1.16}{T} \times 10^4 V_p, \tag{53}$$

where V_p is the anode potential, and T is the absolute cathode temperature.

For $T = 900°$ K,

$$\eta_0 = \eta_0' + 12.9 V_p. \tag{54}$$

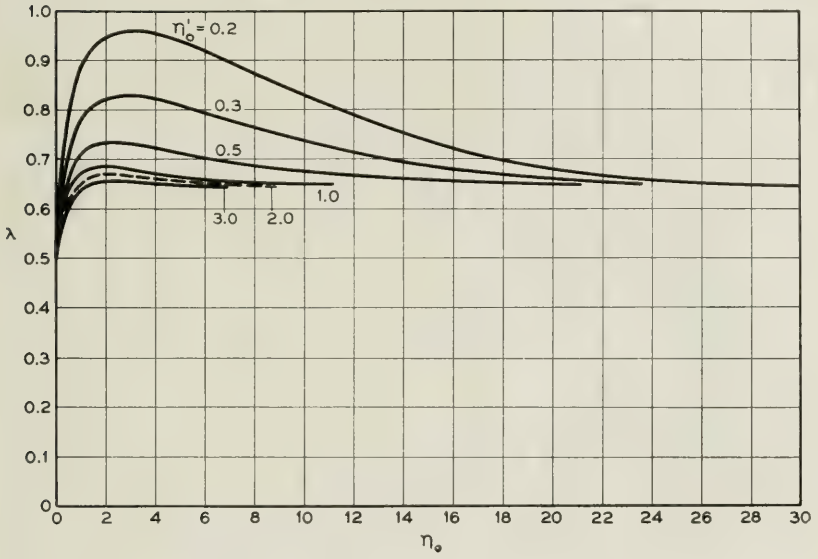


Fig. 2—Effective noise generator voltage of planar diode.

$\overline{E_f^2} = 4kr_p(\lambda T)df$, $\eta_0' = \log_e \frac{I_s}{I_p}$, $\eta_0 = \eta_0' - \frac{e}{kT} V_p$.

For $T = 900°$ K., $\eta_0 = \eta_0' + 12.9 V_p$.

Even for the small space charge condition for which the plate current is eight-tenths of the saturation current ($\eta_0' = 0.2$), the value of η_0 need be greater than about 25 only before λ assumes its limiting value. For a temperature of 900° K., as for oxide coated cathodes, this would require a plate voltage of only two volts. If the plate current were less than eight-tenths of the saturation current for very high plate voltages, then as the plate voltage is reduced, η_0' would increase. For this operating condition λ maintains its limiting value of 0.644 for all except negative values of plate voltages.

The transition between the various effective planar diode plate resistance temperatures is more clearly shown in Fig. 3. In this

figure, the natural logarithm of the ratio of saturation current to the plate current is plotted as a function of the plate voltage for several constant values of the coefficient λ. These curves show for a fixed positive value of plate voltage that as the space charge is decreased toward zero, by a reduction in the ratio of the saturation current to the plate current, the value of λ for moderately large values of space charge increases but little from 0.644, and then for a very low space charge, increases very rapidly to its limiting value given by the shot noise

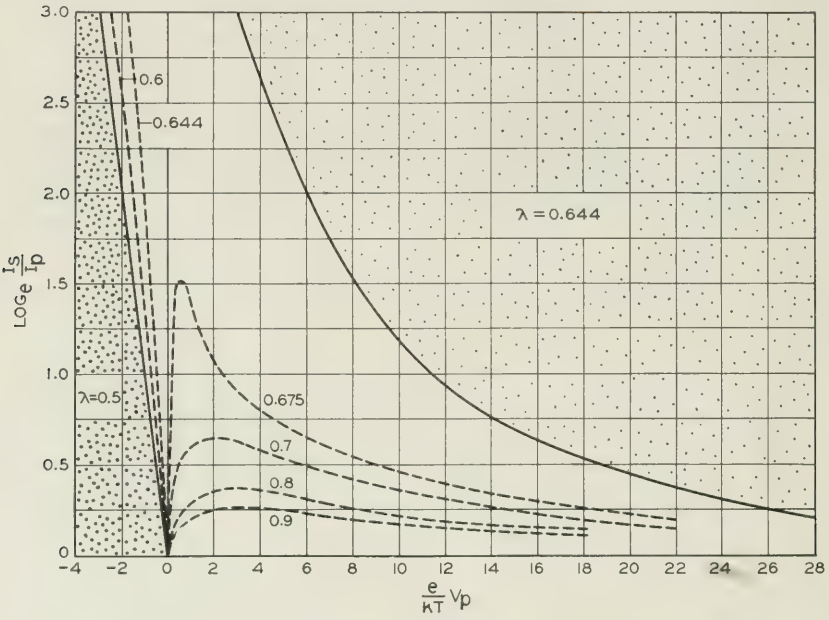


Fig. 3—Modification of effective plate resistance temperature produced by space charge.

$$\overline{E_f^2} = 4kr_p(\lambda T)df.$$

which is represented by the axis of abscissæ. Thus, the value of λ digresses markedly from 0.644 only for the narrow region of operating conditions for which the saturation current is less than 1.25 times the plate current and the plate voltage is less than $28e/kT$ volts. For an oxide coated cathode for which T is 900° K., the effective plate resistance temperature is $0.644T$ for any operating condition for which plate current is less than eight-tenths of the saturation current and the plate voltage is greater than two volts.

For a cylindrical diode, the general method of analysis used in the parallel plane case results in equations which are practically impossible

to solve. The difficulty in these equations arises from the fact that tangential as well as radial initial velocities must be considered in obtaining the total anode current. Since it was shown for the planar diode that the effective temperature of the plate resistance is 0.644 times the cathode temperature for practically all operating conditions, all that is really desired in the cylindrical diode solution is the limiting value of the effective tube temperature. This may be found rather easily from a comparison of the cylindrical diode with the planar tube in the following manner.

For a very large space charge, and a high plate potential the radius of an equipotential surface near the potential minimum will be very nearly equal to that of the cathode. Hence, for these operating conditions, the planar diode equations may be applied to this region of the cylindrical diode. In the planar tube, it was shown that for $\eta_0' > 3$, η_0 had to be of the order of unity to obtain the limiting value of 0.644 for λ. If the space charge and plate potential are sufficiently large in the cylindrical diode, the radius of the equipotential surface for which η_0 is greater than unity will practically be equal to that of the cathode. The cylindrical diode may then be divided into two parts, a planar diode between the cathode and the equipotential surface for which $\eta_0 > 1$, and a cylindrical diode formed from the remainder of the tube. In any diode, the only source of noise energy is the cathode from which the noise power is transferred to the anode and external circuit through the mechanism of the initial electronic velocities. Furthermore, the same total noise power must be transferred across any equipotential surface between the cathode and anode. In the planar portion of the cylindrical diode as described above, the total noise power crossing any equipotential surface was shown to be $2.576kTdf$. This same noise power must be transferred across any other equipotential surface in the cylindrical diode. Hence, the effective plate resistance temperature for the cylindrical electrode tube must also be 0.644 times its cathode temperature. From this line of reasoning, it may be shown that the limiting value of the effective temperature for any shape diode is the same as that for the planar tube with the same cathode temperature.

From the experimental data given in his paper, Pearson definitely recognized that the limiting value of the diode plate resistance temperature should be between 0.59 and 0.65 of that of the cathode.[2] The writer understands that North and Thompson of the R.C.A. in an unpublished paper have obtained the same general result for the effect of space charge upon shot noise in diodes.

In a diode, the tube noise may be expressed equally well and with equal correctness either as a modified shot noise or as a thermal resistance noise. In this paper, the thermal resistance viewpoint was taken for two reasons. First, the coefficient "λ," used in the thermal resistance noise equation

$$\overline{E_f^2} = 4kr_p(\lambda T)df,$$

is practically always a constant equal to 0.644, whereas, the factor, "F," used by Schottky and Spenke in their modified shot noise equation

$$\overline{I_1^2} = 2eF^2I_0df$$

is always a function of the operating condition. That is, for the operating conditions for which λ is a constant, F has the following value:

$$F = \frac{1.39}{\left[\log \frac{I_s}{I_p} - \frac{e}{kT} V_p\right]^{1/2}}.$$

The second reason for the selection of the thermal resistance noise relation is that power from the motion of the atoms in the cathode is actually transferred to the plate electrode and external circuit through the mechanism of the initial electron velocities. Hence, the tube noise in a diode with space charge is very similar to a thermal resistance noise.

Part III—Effect of Transit Time

The analysis, in Part I, while giving the correct results for all operating conditions in the ordinary frequency range, is extremely long and cumbersome. It shows, however, that only the limiting values of the effective temperature of the plate resistance are required for most practical cases, and therefore it points the way to make simplifying assumptions which result in a much shorter analysis, and moreover, which allow the analysis to be extended to frequencies so high that electron transit time phenomena become of importance.

Thus the final noise equation in Part I shows that for moderately high anode potentials and for the usual excess of cathode emission, a very good approximation may be had by a consideration of the current-voltage relations existing in the β-region between potential minimum and anode without the necessity of encumbering the analysis by including the α-region between potential minimum and cathode. Moreover, for a large anode potential, the terminal velocities of the

electrons at the plate are very large in comparison with their initial velocities for practically all of the electrons. This means that the transit time for the various electrons is practically the same for all of them which leave the cathode within a particular very short time interval, even though the initial velocities of the various electrons are statistically distributed among them. It results that the various individual velocities of the electrons in the β-region may be replaced by an average value, which at the potential minimum may be defined as follows:

$$\bar{u} = \frac{\int_{u_c'}^{\infty} u' n(u_c) du_c}{\int_{u_c'}^{\infty} n(u_c) du_c} . \tag{55}$$

Physically, the meaning of this expression is the average velocity of these electrons which cross a plane in the β-region close to the potential minimum in a unit of time. Inasmuch as the unit of time may be taken to be very small, it follows that (55) expresses the effective instantaneous value of the initial velocity which may, and does, fluctuate as time goes on.

On the basis of an equation of the form

$$I = \rho u - \epsilon \frac{\partial^2 V}{\partial t \, \partial x} \tag{56}$$

the planar diode has been extensively investigated by a number of workers and it has been shown [8] that the relation between current and voltage is completely specified as soon as two boundary conditions are given. These may be the initial velocity and acceleration, or they may equally well be the initial velocity and conduction current ρu. However, the analysis based on (56) applies strictly to the case where all of the charge moves with the same velocity and hence contains a certain approximation when electrons are considered whose velocities have a certain dispersion around some mean value. The error will be small until frequencies are considered which are so high that a large proportion of the electrons which left the cathode in a time interval which is very short compared with the period of the high frequency arrive at the anode in a time interval which is not small compared with the high frequency period. Normally this means that the error is small even for frequencies so high that the majority of the electrons require several cycles to make their transit from potential minimum to anode.

It is convenient to write the resulting equations in terms of d-c. and first order a-c. values where the initial values of d-c. velocity and acceleration are given, but initial values of a-c. velocity and conduction current are employed. The first order a-c. relation derived by Llewellyn may be written in the form

$$-\frac{e}{hm}V_1 = \frac{e}{hm\epsilon}I_1A + \frac{e}{hm\epsilon}q_aB + \mu_aC, \tag{57}$$

where q_a and μ_a are the initial values of fluctuation conduction current and velocity, respectively, while A, B and C are defined by:

$$\left.\begin{aligned} A &= \frac{1}{\omega^4}\left[-i\omega^3x + \frac{eI_0}{hm\epsilon}(2 - 2e^{-i\theta} - i\theta - i\theta e^{-i\theta})\right] \\ B &= -\frac{1}{i\omega^3}[a_a(i\theta e^{-1\theta} + e^{-i\theta} - 1) + u_a i\omega(e^{-i\theta} - 1)] \\ C &= \frac{eI_0}{hm\epsilon\omega^2}[i\theta e^{-i\theta} + e^{-i\theta} - 1] \end{aligned}\right\}, \tag{58}$$

in which η is the transit angle, $\omega\tau$, the transit time being τ, and I_0 is the d-c. current.

In the application of these relations to noise analysis, the initial values of velocity, acceleration, and conduction current must be taken at a point in the β-region beyond the potential minimum, but just as close to it as possible without encountering conditions where electrons may be moving toward the cathode, for the equations apply only to cases where the electrons are moving in one direction only. The initial point is, however, located so near to the potential minimum that the d-c. acceleration in (58) may be taken as zero. When this is the case, it may be shown that the initial conduction current is equal to the total current. In other words, the initial value of displacement current is zero. Under such conditions (57) and (58) reduce to the following expression for the a-c. anode potential in terms of the a-c. component of current and initial velocity:

$$V = \frac{I_1}{\omega^4\epsilon}\left[\frac{eI_0}{hm\epsilon}\left(\frac{i\theta}{6} + i\theta + 2e^{-i\theta} + i\theta e^{-i\theta} - 2\right) + \omega^2u_a(i\theta + e^{-i\theta} - 1)\right] + \frac{\mu_aI_0}{\omega^2\epsilon}[i\theta e^{-i\theta} + e^{-i\theta} - 1]. \tag{59}$$

The term multiplying the a-c. current I_1 in the above equation is the internal high-frequency impedance z of the planar diode. The last term may therefore be identified with an internal emf. When the initial velocity μ_a is expressed in terms of the fluctuation of electron

velocity, the term gives the equivalent noise generator, E. Thus

$$E = \frac{\mu_a I_0}{\omega^2 \epsilon} (i\theta e^{-i\theta} + e^{-i\theta} - 1) \tag{60}$$

and the mean-square value of the noise emf. (at a frequency ω) is given by:

$$\overline{E^2} = \frac{\overline{\mu_a{}^2} I_0{}^2}{\omega^4 \epsilon^2} |i\theta^{-i\theta} + e^{-i\theta} - 1|^2. \tag{61}$$

The problem is now reduced to finding the mean square value of initial velocity fluctuation, $\overline{\mu_a{}^2}$, which corresponds to electrons crossing the potential minimum. This may be done by going to (55) which gives the effective value of the instantaneous initial velocity and separating all quantities, including the lower integration limits into d-c. and a-c. components. Thus

$$\left.\begin{aligned} n(u_c) &= n_0(u_c) + \delta(u_c) \\ u_c' &= \overline{u_c'} + \delta u_c' \\ u' &= \overline{u'} + \delta u' \\ \bar{u} &= u_a + \mu_a \end{aligned}\right\}. \tag{62}$$

The result may be expanded in series form and products of the δ's may be disregarded inasmuch as the a-c. components are small in comparison with the d-c. The indicated operations have as a result

$$u_a = \sqrt{\frac{\pi k T}{2hm}} \tag{63}$$

and

$$\mu_a = \frac{e}{I_0} \int_{\bar{u}_c}^{\infty} (u' - u_a)\delta(u_c) du_c. \tag{64}$$

The Fourier analysis may be applied to this in the way outlined in connection with (37) and (41) in Part I and gives the mean-square value of velocity fluctuation corresponding to a frequency interval df as follows:

$$\overline{\mu_f{}^2} = \frac{2e^2}{I_0{}^2} df \int_{\bar{u}_c}^{\infty} (u' - u_a)^2 u_c(u_c) du_c = \frac{4ekT}{I_0 hm} df \left(1 - \frac{\pi}{4}\right). \tag{65}$$

This may be substituted in (62) giving for the effective noise emf. in the frequency range df

$$\overline{E_f{}^2} = 4kTdf \left[\frac{eI_0\tau^4}{hm\epsilon^2}\right] \left[1 - \frac{\pi}{4}\right] \left[\frac{1}{\theta^4}\right] \times \tfrac{1}{4}[\theta^2 + 2 - 2(\cos\theta + \theta \sin\theta)]. \tag{66}$$

The initial average velocity is small so that the low-frequency plate impedance may be written

$$r_p = \frac{eI_0\tau^4}{12hm\epsilon^2}. \tag{67}$$

Thus for any transit angle, the mean square noise generator voltage is given by

$$\left.\begin{aligned}\overline{E_f^2} &= 12\left(1 - \frac{\pi}{4}\right) kr_pTdf\left\{\frac{4}{\theta^4}[2 + \theta^2 - 2(\cos\theta + \theta\sin\theta)]\right\} \\ &= 4Sk(0.644T)r_pdf\end{aligned}\right\}. \tag{68}$$

where

$$S = \frac{4}{\theta^4}\left[2 + \theta^2 - 2(\cos\theta + \theta\sin\theta)\right]$$

For low transit angles, this expression reduces to

$$\overline{E_f^2} = 12\left(1 - \frac{\pi}{4}\right) kr_pTdf, \tag{69}$$

which is precisely the limiting value obtained by the much longer, but more rigorous analysis.

It must be understood that (68) is an approximation since the transit time effect in the region between cathode and potential minimum was entirely neglected, and because the validity of the average velocity concept does fail at the very high frequencies.

Some knowledge of the extent of the operating conditions for which the above equations are good approximations may be obtained from the d-c. current-voltage relation. For the boundary conditions assumed, the low frequency current equation derived from the general solution given by Llewellyn reduces to

$$I = \frac{-2.33(V - V')^{3/2}}{10^6(x - x')^2}\left[1 + 2.66\sqrt{\frac{kT}{e(V' - V)}}\right]. \tag{70}$$

This equation was shown by Langmuir to be a very good approximation for the plate current for most operating conditions and fails only for very low values of plate voltages. Thus, it may be concluded that (68) is a good approximation for all operating conditions except for very low plate voltages and a small space charge.

The plot of (68) given in Fig. 4 shows that the magnitude of the mean square noise generator voltage decreases by five per cent only for transit angles as large as one radian.

The effect of transit time on the pure shot noise for a low space charge density and a high plate voltage may be obtained quite readily from (57) and (58). Since for a very small space charge, I_0 and u_a are small, and a_a large, the equations then reduce to the following expression:

$$-V_1 = \frac{I_1}{i\omega\epsilon}\frac{a_a\tau^2}{2} + \frac{q_a}{i\omega\epsilon}a_a\tau^2\left[\frac{1}{-\theta^2}(i\theta e^{-i\theta} + e^{-i\theta} - 1)\right]. \qquad (71)$$

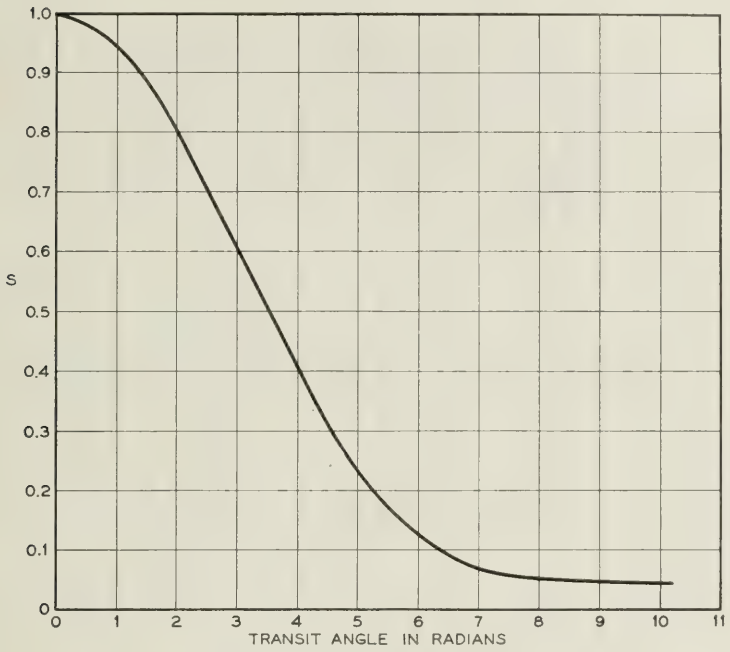


Fig. 4—Effect of transit time on both thermal and shot tube noise.

$$\overline{E_f^2} = 4Sk(0.644T)r_p df,$$
$$\overline{I_f^2} = 2eSI_0 df.$$

For these operating conditions, the transit time in terms of the d-c. acceleration and the electrode spacing is given by

$$x = \frac{a_a\tau^2}{2}.$$

In terms of the external circuit impedance Z_f

$$V_1 = I_1 Z_f, \qquad (72)$$

so that (71) and (72) combine to give

$$I_1 = -\frac{q_a}{1 + \frac{Z_f i\omega\epsilon}{x}} \left[\frac{2}{i\theta^2} (i\theta e^{-i\theta} + e^{-i\theta} - 1) \right]. \tag{73}$$

The mean square shot noise current is thus given by

$$\overline{I_1^2} = \frac{\overline{q_a^2}}{\left| 1 + \frac{Z_f i\omega\epsilon}{x} \right|^2} \left| \frac{2}{\theta^2} (i\theta e^{-i\theta} + e^{-i\theta} - 1) \right|^2. \tag{74}$$

The value of the mean square a-c. conduction current at the cathode to be substituted in the above equation may be derived as follows:

The total current emitted from the filament was defined as

$$I = e\int_0^\infty n(u_c)du_c = e\int_0^\infty n_0(u_c)du_c + e\int_0^\infty \delta(u_c)du_c. \tag{75}$$

Hence

$$q_a = e\int_0^\infty \delta(u_c)du_c. \tag{76}$$

From (37) and (41), the contribution to the mean square of this conduction current from the frequencies between f and $f + df$ is

$$\overline{q_a^2} = 2e^2 df \int_0^\infty n_0(u_c)du_c = 2eI_0 df. \tag{77}$$

With this result, the effect of transit time on the shot noise current is given by

$$\overline{I_f^2} = \frac{2eI_0 df}{\left| 1 + \frac{Z_f i\omega\epsilon}{x} \right|^2} \left\{ \frac{4}{\theta^2} [2 + \theta^2 - 2(\cos\theta + \theta \sin\theta)] \right\}, \tag{78}$$

where Z_f is the impedance of the external circuit at the frequency f and $x/i\omega\epsilon$ is the capacitive reactance of the diode at the same frequency. Thus the shot noise current is modified by transit time in precisely the same manner as the noise generator voltage for the thermal tube noise.

The effect of transit time upon the shot noise, as indicated in (78), is identical with that obtained by Spenke for the same operating condition of low space charge and high anode potential.[4] Spenke derives this result through a clever application of a Fourier Series in which account was taken of the effect of transit time upon the wave shape of the current induced in the anode by the electron moving from

cathode to the plate. The advantage of the method of average velocities used in this paper is that the effect of transit time in both the thermal tube noise and the shot noise may be found.

It is noteworthy that Ballantine in 1928 derived an expression for the effect of transit time upon the pure shot noise which is identical to that obtained in this paper.[9]

In conclusion, the writer wishes to express his appreciation to F. B. Llewellyn whose supervision and numerous suggestions made possible this paper.

References

1. F. B. Llewellyn, "A Study of Noise in Vacuum Tubes and Attached Circuits," *Proc. I.R.E.*, vol. 18, pp. 243–265 (1930).
2. G. L. Pearson, "Shot Effect and Thermal Agitation in a Space Charge Limited Current," *Physics*, vol. 6, pp. 6–9 (1935).
3. W. Schottky and E. Spenke, "Die Raumladungsschwächung des Schroteffektes," *Wissenschaftliche Veröffentlichungen aus den Siemens-Werken*, vol. 16, pp. 1–41, Aug. 6, 1937.
4. E. Spenke, "Die Frequenzabhangigkiet der Schroteffektes," *Wissenschaftliche Veröffentlichungen aus den Siemens-Werken*, vol. 16, pp. 127–136, Oct. 8, 1937.
5. L. H. Germer, "Distribution of Initial Velocities Among Thermonic Electrons," *Phys. Rev.*, vol. 25, 795 (1925).
6. T. C. Fry, "Thermonic Current Between Parallel Plane Electrodes," *Phys. Rev.*, vol. 17, 441 (1921).
7. I. Langmuir, "Effect of Space Charge and Initial Velocities," *Phys. Rev.*, vol. 21, 419 (1923).
8. F. B. Llewellyn, "Operation of Ultra High Frequency Vacuum Tubes," *B.S.T.J.*, Oct. 1935.
9. S. Ballantine, "Schrot-Effect in High Frequency Circuits," *J.F.I.*, vol. 206, 159 (Aug. 1928).
10. T. C. Fry, "The Theory of the Schroteffekt," *J.F.I.*, vol. 199, No. 2 (Feb. 1925).

Fundamentals of Teletypewriters Used in the Bell System

By E. F. WATSON

During the past few years the use of teletypewriters has become quite extensive in the Bell System. Simpler and cheaper machines have recently been made available for meeting the simpler service requirements and attachments have been designed to provide additional features for meeting more complex service requirements. This article discusses the fundamental principles and various features of the teletypewriter machines now in common use and explains the more important factors which have been controlling in their development.

WITH the growth of Teletypewriter Exchange Service and the general increase in the use of teletypewriters in private line services of various types, questions frequently asked are: How do teletypewriters operate? What is the "start-stop" system? Why is it used? What is a regenerative repeater?

This article will attempt to answer some of these questions and explain also the fundamental principles and features of teletypewriters and their auxiliary arrangements as now employed in the Bell System. These have been developed to meet the needs of customers for a typed or similar record form of communication and at the same time be suitable for operation in connection with the Bell System plant.

Code

For economical transmission over long distances it is fundamental that only a single wire or transmission channel be required to carry the signals. Furthermore, long experience with manual telegraphy on land lines has proved that reliable and efficient operation is secured by using not more than two conditions on the line, such as current and no current or positive impulses and negative impulses, as contrasted with the use of three or more conditions, or current values. The entire telegraph plant of the Bell System as well as practically all other land line telegraph systems have been built on this two-condition basis.

The familiar Morse code uses sequences of dots and dashes to represent the different characters of the alphabet and meet the above conditions. This code is not well adapted for teletypewriter control, however, since the signals for different characters vary widely in the time they require, from a single dot for the letter E to a combination of

several dots and dashes for some of the less frequently used letters or numerals.

For machine operation it has thus far appeared desirable in order to obtain simplest mechanisms and to obtain maximum operating speeds with low line signaling frequencies, to have the signals for the different characters of uniform length, that is, each contain the same number of time units. This condition is met by the five-unit code where each character is identified by the impulses in five units of time, and this is the code normally employed in Bell System teletypewriters. Each of the five units of this code may be either positive or negative, current or no current, or either of two values of current, and the permutations provided are 2^5 or 32. These are sufficient for the 26 letters of the alphabet, a space, carriage return and paper feed signals as well as case shifting signals to bring another set of characters into action so as to include numerals and punctuation marks. A chart of this code as used in Teletypewriter Exchange Service (TWX), is shown below.

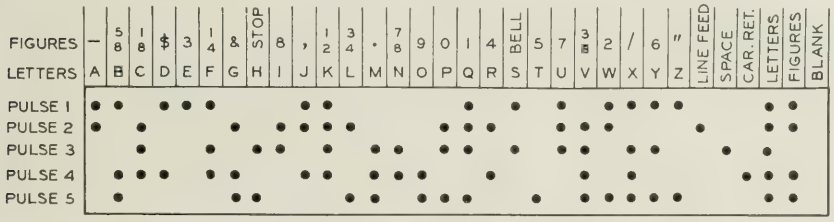

FIGURES — 5/8 1/8 $ 3 1/4 & STOP 8 , 1/2 3/4 . 7/8 9 0 1 4 BELL 5 7 3/8 2 / 6 " LINE FEED SPACE CAR. RET. LETTERS FIGURES BLANK
LETTERS A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
PULSE 1
PULSE 2
PULSE 3
PULSE 4
PULSE 5

Fig. 1—Chart of five-unit TWX code.

The keyboard used for sending this code is shown below.

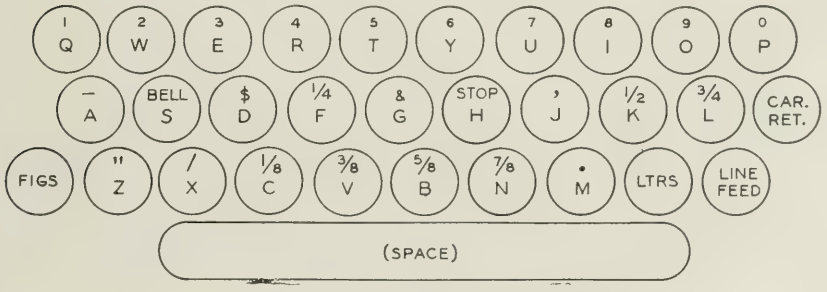


Fig. 2—Chart of TWX keyboard.

It will be noted that this keyboard is similar to the ordinary typewriter keyboard except that there are only three rows of keys instead of four as in the typewriter. In the typewriter keyboard, the lower three rows of keys are used ordinarily for small letters but when a shift

key is also operated they type the corresponding capital letters. The fourth or top row of keys carries the numerals and certain punctuation marks. The teletypewriter types capital letters but not small letters so that by using a shift or *Figures* key the upper case position of the letter keys is available for the usual punctuation marks and numerals. Thus only three rows of keys are required on the teletypewriter keyboard. The operation of the *Figures* key sends a signal causing the receiving machine to shift to upper case so that numerals and punctuation marks will be printed until a *Letters* or *Space* signal is sent which restores the machine to lower case.

Start-Stop System

For transmitting the signals of the five-unit code over a telegraph line, it is necessary to have some system of timing so that each of the five impulses may be properly received, identified and interpreted at each receiving station. The start-stop system is used for this purpose. One arrangement of this system using segmented distributors with revolving brushes, is illustrated in Fig. 3.

In this system both sending and receiving brush arms are normally at rest but are maintained under constant torque, tending to rotate them in the direction of the arrows, by constantly running motors driving them through friction clutches. Normally the line circuit is closed and carries current. When a key of the keyboard is operated to send a signal, the start magnet of the sending distributor is energised releasing the sending brush arm and allowing it to rotate. As this brush passes from the stop to the start segment, the line circuit is opened and this open signal transmitted to the receiving station where it causes energization of a start magnet which releases the receiving brush and allows it to rotate.

Both sending and receiving brush arms rotate at approximately the same speeds since they are driven from motors running at approximately the same speeds. These motors are either small synchronous motors driven from constant frequency commercial 60-cycle 110-volt power supply or by commutator type motors, equipped with centrifugal governors to hold them at approximately a constant speed, for use on other commercial a-c. or d-c. supplies.

Now as the sending brush arm sweeps over the sending face, the impulses of the five-unit code, as set up by the particular key depressed, will be transmitted over the line as shown in Fig. 4 for the letter A, and through the action of the rotating receiving brush, Nos. 1 and 2 current impulses will cause the energization of Nos. 1 and 2 selecting

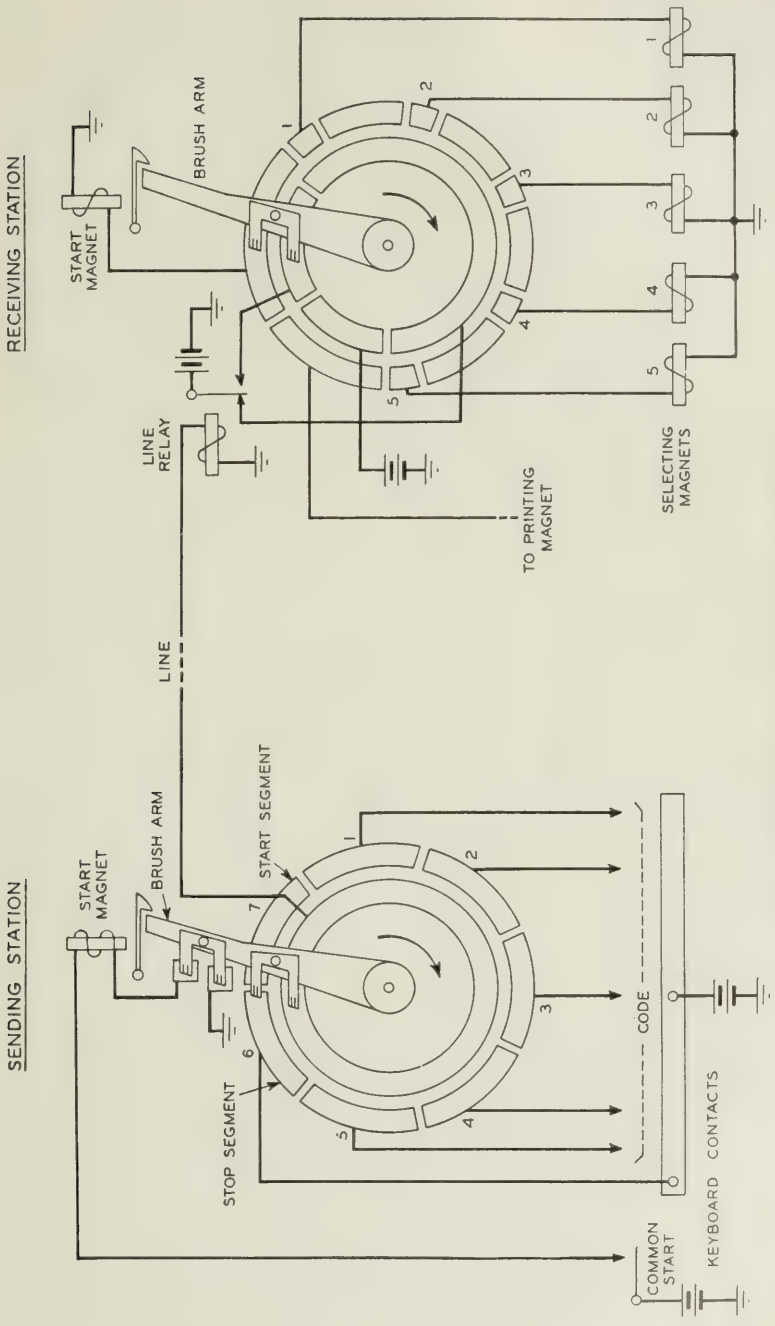


Fig. 3—Simplified diagram of start-stop system.

magnets while Nos. 3, 4 and 5 no-current impulses will not energize selecting magnets 3, 4 and 5.

Similarly other signals for other characters may be sent and received as permutations on the selecting magnets whose operation may in turn control the selection of an individual type bar to be operated to type the desired character.

While brush and segment distributors are shown above for purposes of illustration, in modern machines these are replaced by simple mechanical devices, sending contacts and a single receiving magnet, which function in the same manner but are cheaper, more reliable and easier to maintain.

Advantages of the start-stop system over other systems of timing include its simplicity, the fact that highly accurate speed regulation of the motors is not required, that to start a station it is only necessary to turn on the power, and that lag in the line signals due to time for propagation is automatically compensated for. The lag compensation

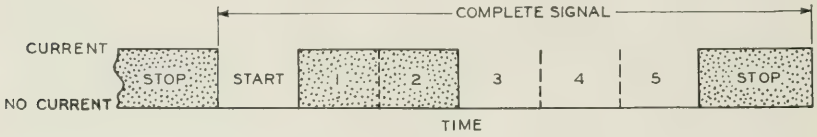


Fig. 4—Chart of line signal for letter A.

is automatic because the receiving distributor does not start until the start signal has been propagated and received and then the other signals always follow in a fixed definite time relation. This makes it possible to connect any number of stations to a single circuit and have them intercommunicating; that is, signals sent from any station will be received and printed at all other stations without requiring readjustments, regardless of the distances or circuit complications due to repeaters and the like intervening.

To illustrate one of these points further, it may be mentioned that with the start-stop system, if other distortions were not present, it would be theoretically possible to secure perfect operation without errors between two ideal distributors one of which was running either faster or slower than the other by as much as 7 or 8 per cent. This is because a correction is automatically made after each character is transmitted. Since some distortion of the signals is usually experienced in transmission, in order to maintain a large tolerance for such distortions, it is the usual practice to attempt to maintain the speed of each distributor within $\pm$ .75 per cent although a much

larger variation may be tolerated without causing errors except when it occurs simultaneously with an abnormally large line distortion.

In contrast with this, in a synchronous system such as employed in the usual "Multiplex Printer System," it is necessary that the speeds of distributors be very accurately maintained or errors in transmission will result. This is accomplished by controlling the driving motors from very accurate timing sources such as tuning forks, then testing the speed of the receiving distributor two or more times every revolution from the signals transmitted over the line from the sending distributor, and automatically correcting this speed as required. If it were not for this last mentioned correction in order to receive signals correctly on a two-channel multiplex system for a period of 15 minutes at a speed of 60 words per minute per channel, the receiving distributor speed would have to be held accurate within $\pm$ .002 per cent. In other words, in 15 minutes the receiving distributor would make about 5400 revolutions (60 words per minute $\times$ 6 revolutions per word $\times$ 15 minutes). At the end of the 15 minutes, if it were one-tenth of a revolution ahead or behind its correct position, a No. 3 pulse, for example, would be received on a No. 4 or No. 2 segment resulting in an error. Thus the limit would be that the receiving distributor could not run fast or slow by as much as one-tenth revolution in 5400 revolutions or roughly 2 parts in 100,000.

Tolerance for Distorted Line Signals

During transmission over long lines, telegraph signals may become badly distorted; that is, the dots and dashes may be considerably shortened or lengthened from their correct values. It is essential that the receiving distributors be capable of receiving and interpreting these signals without error. To accomplish this the receiving distributors are arranged so that they are sensitive for the reception of the selecting impulses only for a very short time at the middle of each impulse. The exact location of this sensitive period with relation to the incoming signals is adjustable in each receiving distributor so that it may have maximum tolerance for receiving distorted signals. This situation is illustrated in the diagram, Fig. 5, which shows a receiving distributor with the received signals under various conditions developed, and the sensitive points for the reception of the impulses indicated by small arrows.

It will be noted from this chart that the signals may be very badly distorted (theoretically up to nearly 50 per cent of a pulse length at either the front or rear end of a current pulse), without causing imperfect reception.

In the figure it will be noted that trace (a) shows a perfect signal with the arrows denoting the sensitive points for receiving each pulse located directly at the center of each of the pulses 1 to 5. This is a normal adjustment for a receiving distributor which, without change of this adjustment, must be able to tolerate the distortions of various kinds

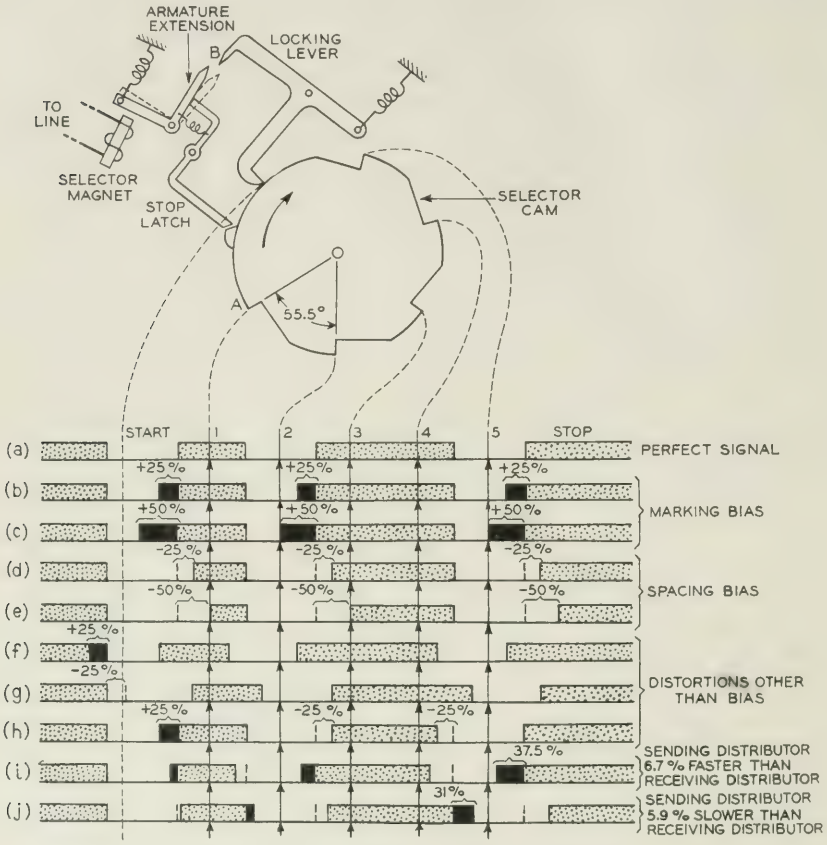


Fig. 5—Effect of distorted signals on reception.

experienced in service without failure to receive and properly identify each pulse as a No. 1, 2, 3, 4 or 5 pulse. The other traces illustrate certain types of distortions which may be experienced and the conditions existing for their proper reception with this adjustment. These traces will now be explained on a purely theoretical basis assuming an ideal receiving machine without mechanical or other imperfections.

Trace (b) shows the conditions in case of 25 per cent "marking bias" in the received signals; that is, each marking pulse has been lengthened

by 25 per cent of one pulse length. It will be noted that under these conditions each pulse will still be properly received and identified.

Trace (c) illustrates 50 per cent marking bias in the received signal, and at this point it will be noted that the No. 3 and stop pulses have been so elongated that they are just on the verge of being erroneously received and identified also as No. 2 and No. 5 pulses. This then is a theoretical limit for proper operation with marking bias without a readjustment of the receiving distributor.

Similarly traces (d) and (e) illustrate respectively the conditions when the received signals have 25 per cent and 50 per cent "spacing bias," that is each marking impulse has been shortened by this percentage of one pulse length. It will be noted that 25 per cent spacing bias can be easily tolerated but that with 50 per cent spacing bias the No. 1 and No. 3 pulses are on the verge of failure to be recorded. This then is a theoretical limit for spacing bias in the signals under this adjustment.

Traces (f) and (g) show the effect of 25 per cent marking and spacing distortions respectively on the rear end of the stop or front end of the start impulse, all other pulses remaining undistorted.

Trace (h) shows the effect of distortions on the selecting impulses alone. By combining this trace with traces (f) and (g) it will be seen that with 25 per cent distortion of the start pulse, 25 per cent distortion of the same sign is the limit for distortion on the front end of marking impulses, or 25 per cent distortion of opposite sign on the rear end of marking impulses. Thus for distortions other than bias, which are apt to affect both start and selecting impulses in the same signal, $\pm$ 25 per cent is the theoretical limit of allowable distortion.

Traces (i) and (j) show the effects of speed inaccuracies. From these it will be seen that theoretically the sending distributor could be about 8.9 per cent faster or 9.5 per cent slower than the receiving distributor before errors would be experienced.

In practical machines there are, of course, inaccuracies due to tolerances of manufacture, and other departures from the ideal so that the above mentioned theoretical limits are not reached. However, all machines used in the Bell System are required to tolerate a lengthening or shortening of the front end of any current impulse of at least 40 per cent of its length and with the same adjustment a lengthening or shortening of the rear portion of any current impulse of at least 35 per cent with the start pulse undistorted. Since bias is nearly always present to some degree in the received signals, and since as interpreted by the receiving distributor it affects only the front end of the current impulses as illustrated in traces (b), (c), (d) and (e) of Fig. 5, the

distributors are usually adjusted for maximum tolerance of front end distortions and then have ample tolerance for such distortions of the rear ends of the current impulses as are experienced under service conditions.

Regenerative Repeaters

As circuits become longer and more complex, eventually a point is reached beyond which signal distortion becomes so great that the signals cannot be reliably received without error. To overcome this

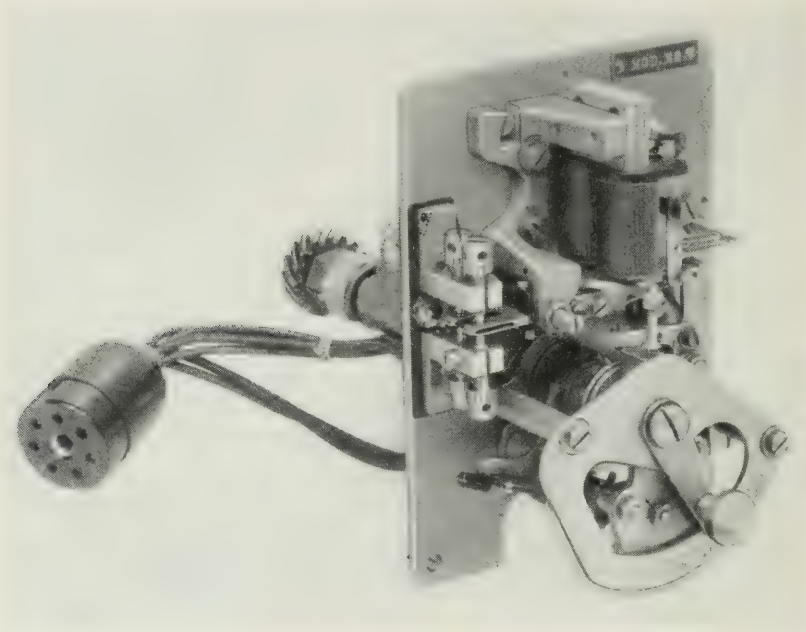


Fig. 6—Regenerator unit.

limitation a device known as a regenerative repeater may be inserted in the line at this point. It has a receiving mechanism similar in principle to the receiving distributor of a teletypewriter and will accurately receive and interpret any signals which a teletypewriter would accurately record. This receiving mechanism is interconnected with a retransmitting mechanism, or sending distributor, which retransmits the signals reshaped and reformed so as to be substantially free from distortion. In its latest form a one-way regenerative repeater consists of a receiving magnet and a set of transmitting contacts interconnected by some relatively simple mechanical parts driven from

a motor. A photograph of such a recently developed regenerative repeater unit is reproduced in Fig. 6.

By means of these regenerative repeaters reliable teletypewriter service may be extended to any desired distances so long as the signals in any one regenerative repeater section are not too badly distorted to permit reliable operation of that section alone. Several regenerative repeaters may be operated in tandem on a single very long circuit if required and in fact a number of very difficult long circuits are operating satisfactorily under these conditions at the present time. A point worthy of note and which has not previously been mentioned is that the stop impulse in the code adopted for Bell System apparatus is slightly longer than the other impulses, which facilitates the use of regenerative repeaters in tandem without requiring complex speed control arrangements.

General Features of Teletypewriters

Teletypewriters are widely used for high speed written communications. Generally speaking, written communications are desired for purposes of accuracy. Therefore, high speed, accuracy and reliability are basic requirements for teletypewriter service.

In choosing an operating speed at which distributors of teletypewriters are to be set, several factors must be considered. These are the capabilities of the mechanisms of the machines, the average capabilities of operators for continuous sending at high speeds, the commercial need for high speeds, and the capabilities of the line circuits for transmitting the signals reliably over long periods without excessive distortions or excessive attention for maintenance and adjustment. A satisfactory compromise among these different factors seems at the present time to be about 60 words per minute, or 368 machine operations per minute, which is the speed usually employed in the Bell System. The machines themselves may be arranged and adjusted to be capable of higher speeds up to about 75 words per minute, and it may be that, in the future, service at these higher speeds will be justified under certain circumstances.

Accuracy and freedom from breakdown troubles are necessarily inter-related and both required to a very high degree for machines handling important written communications over long distances. To give some idea of the severity of these requirements, we have found from long experience that to produce a good machine we can not be satisfied in our laboratory tests unless the machine is capable of typing at least 1,000,000 consecutive words (6,000,000 operations) without

error or trouble of any kind, and without requiring service attention of any kind other than normal replacement of paper and inking ribbons.

For rendering service economically with teletypewriters on subscribers' premises, an important requirement in order that the expense of maintenance be not prohibitive is that the machine should not require maintenance attention except at very infrequent intervals. Bell System machines are designed to require routine maintenance attention not oftener than once in two months where the machine is used continuously over periods of eight hours each day. To accomplish this the problem of lubrication has required very careful attention. It has necessitated the provision of oil reservoirs in certain places and the careful selection and specification of oils and greases. Another feature making for economical maintenance is interchangeable parts. In other words, if a part breaks or wears, it is replaceable by another part of the same type without requiring fitting and usually without readjustment.

At times customers wish to use teletypewriters on tables especially designed and arranged to suit the convenience of their offices. For this reason teletypewriters are designed as far as feasible to be self-contained units which can be mounted on any desk or table.

All present Bell System teletypewriters employ the start-stop system of synchronizing and are well adapted for the connection of any number of machines to one circuit with facilities for rapid to and fro intercommunication among the various stations. To permit optimum control of intercommunication and interruption of the sending station when desired, a device known as the "break lock" is incorporated in many machines. This device, together with a "break" key located on each machine, provides facilities whereby any station may interrupt a station which is sending, take control of the circuit and send. The operation of the "break" key opens the line transmitting a signal which causes the "break lock" device to function at the station which is sending and automatically stop any further sending from that station until the device is manually restored. This device is very important in the case of transmission from a perforated tape, which is described later.

Motor control devices are of importance for stations which are not in continuous use but which may wish to receive messages from time to time from distant stations without requiring an attendant to turn on the machine. Such devices are used both in private line and in TWX services. In the case of a private line it is often desired to have the machine normally idle with the motor stopped but so arranged that,

when a distant station wishes to send a message, a signal may be sent which will automatically start the motor and condition the receiving machine so that it will properly record the message and then have its motor automatically stopped again at the end of the communication. Various devices are available for this purpose, some operating over the regular signaling circuit and others requiring a separate circuit. Similarly, in the case of TWX service, stations may, if desired, be equipped for unattended service so that, if the station is called and no attendant is present, the teletypewriter motor may be started remotely by the switchboard operator and the station conditioned to record the incoming message at the termination of which the motor can again be stopped by the switchboard operator.

Signal bells are usually provided on the machines so that, if it is desired to call an attendant to a working machine or to call attention to a specially important message being received, the bell can be rung by signals sent over the circuit.

A general feature incorporated in the design of all modern machines, and one which is not often appreciated, is the so-called "overlap." This feature makes high speed possible by overlapping the selecting and printing parts of the receiving operation. In other words it provides for the typing of one character to take place simultaneously with the reception of the selecting impulses for the next character.

Features of Page Teletypewriters

Page teletypewriters have been built in several different forms, notably with a moving paper carriage or a stationary paper carriage and with a typewheel or with type bars for printing. An early design employed a moving paper carriage and a typewheel, with an ink roller for inking the characters on the wheel. With this design it was impractical to make satisfactory carbon copies, the printed record was unevenly inked, and much trouble was experienced due to side printing, that is, unwanted printing of portions of letters adjacent to the desired letter on the typewheel. Furthermore, considerable trouble was had in properly feeding paper from a paper roll through the moving paper carriage.

To eliminate these limitations and troubles it was decided that for general service in the Bell System a new machine should be designed to be capable of making as many carbon copies as a typewriter and that it should use type bars and have a stationary paper carriage. This sort of machine was new in the art and required extensive development work to produce a satisfactory commercial design because of the in-

herent difficulties of moving an automatically operated basket of typebars back and forth in front of the stationary paper. The present standard No. 15 teletypewriter was the ultimate result of this work and has proved very satisfactory in general service over a number of years. It employs a typewriter ribbon for inking, has the paper roll inside the machine cover and makes very satisfactory carbon copies with various types of paper supply without being subject to the paper feed, inking and side print troubles previously experienced.

This machine has also lent itself to meeting later demands from business houses for typing either single or duplicate copies on special printed forms as commonly used in modern business practice. By equipping the platen with sprocket teeth and having feeding perforations along the edges of the forms, all copies of these forms are automatically held in perfect registration during typing at all stations connected to the circuit. In connection with the rapid handling of these forms a further requirement for automatic tabulation has been met by providing a tabulating device which on the transmission of a certain signal causes all carriages to move over rapidly to any predetermined position on the form and stop there for the typing of letters or figures in columns perfectly aligned. This device greatly facilitates the rapid transmission and reproduction of orders and the like on organized printed forms.

With the advent of TWX service a new situation arose in which many of the machines were only infrequently used and then for very short periods to make a single copy only. To render this service economically it seemed desirable to have a less expensive machine and since narrower capabilities were required this seemed entirely feasible. Accordingly a new machine known as the No. 26 teletypewriter has been developed primarily to print a single satisfactory copy although one carbon copy can be made if desired. To obtain low first cost this machine has a moving paper carriage and to secure a satisfactory printed record it employs ribbon inking and a typewheel arrangement which is a sort of cross between conventional typebar and typewheel designs. This typewheel is an assembly employing a small individual type pallet for each separate character. In the process of printing a character, a striking arm somewhat like the shank of a typebar comes forward and forces the individual type pallet against the ribbon to make an impression on the paper. The typewheel is rotated to different positions to select the different characters to be typed. In this way satisfactory inking and a clear cut impression without side print is obtained, which compares favorably with the record obtained

on a typebar machine or typewriter. The entire machine costs appreciably less than the more comprehensive No. 15 machine. The No. 26 machine is illustrated in Fig. 7.

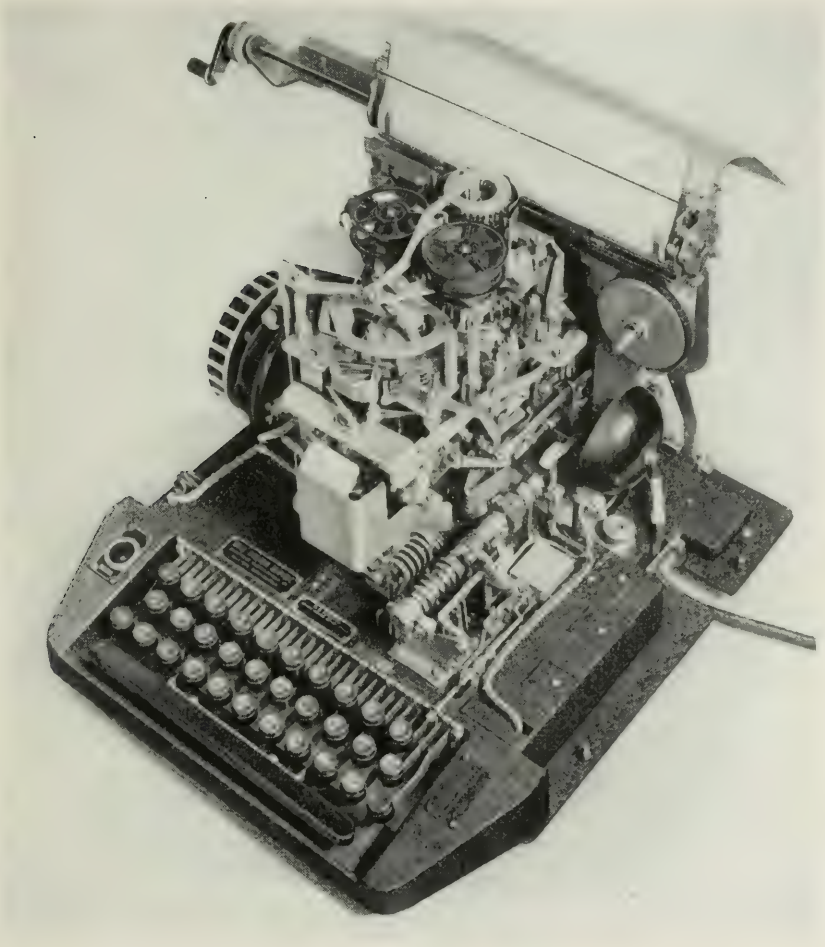


Fig. 7—No. 26 teletypewriter.

Features of Tape Teletypewriters

In the case of tape teletypewriters it is also necessary to have a clean printed record and occasionally there is a need for carbon copies. Accordingly, the tape machine standard for the Bell System is a typebar machine using an inking ribbon and known as the No. 14 teletypewriter. It is illustrated in Fig. 8.

A feature worthy of note is that with this machine typing always occurs at the same point introducing a problem in connection with platen wear. If the platen were fed by the usual ratchet in, say, 36 steps per revolution, there would be heavy wear concentrated at these 36 points and the platen would require frequent replacement to pre-

Fig. 8—No. 14 teletypewriter.

serve good printing. To avoid this, the platen is fed through differential gearing so that on a second revolution the typing comes in a different spot from that of the first revolution; thus the wear is uniformly distributed over the entire circumference.

One carbon copy can be made by leading tapes through the machine from two rolls of record paper and one roll of carbon paper. Two carbon copies can be made in a similar way if desired.

The tapes employed may be either gummed on the back for convenient pasting on blanks for filing or may be plain paper tapes if the records are of temporary interest only. Also cellophane or similar transparent tape may be used if it is desired to project the record on a screen. A tape out signal is provided on the machine so that when a roll of tape becomes nearly exhausted a bell will ring continuously to give warning of this fact. Where a bell is not desired, the last few feet of tape on the roll are painted red to give similar warning.

If desired, this tape printing machine may be used on the same circuit with page printing machines such as the No. 15 teletypewriter, and when so employed is usually equipped with an "end of line indicator" to warn the operator of the approach of the end of the line in the page machine, so that suitable signals may be sent for starting a new line.

Features of TWX Switchboard Operators' Teletypewriters

Such machines must be small in size to permit their use in a switchboard position, quiet in operation to permit their use in the same room with a telephone switchboard and must be capable of working with any machine employed in the TWX system.

To meet these requirements the standard No. 14 tape teletypewriter has been modified in several important respects as follows:

1. It has been provided with a specially designed enclosing cover which reduces the machine noise radiated by at least 5 db more than standard covers.

2. The machine is tilted so as to raise the keyboard and permit the operator to assume a more elevated position nearer the switchboard jack field.

3. It is equippent with an end of line indicator mechanism and lamp to warn of the approach of the end of a line when sending to a page teletypewriter station so that the proper signals may be sent to start a new line.

4. The usual tape feeding mechanism which pulls the tape past the typing point and obscures some of the typed message is replaced by a so-called "push feed" mechanism which acts ahead of the typing point and makes the typed message more fully visible.

5. Many of the operators' machines are provided with specially arranged power supply and governing circuits so that their motors normally run from 115 volt a-c. commercial supply but in case of a power failure can be quickly switched to run from the 130 volt d-c. telegraph battery.

Features of Monitoring Teletypewriters

In connection with private wire teletypewriter service it has been found very desirable to have so-called monitoring teletypewriters in the repeater offices to facilitate testing between offices and with the subscriber stations. These machines must be adaptable to work with any subscriber's machine and to be usable for making test measurements on circuits.

The No. 14 tape teletypewriter has also been adapted to this service. It may be equipped with an end of line indicator to facilitate communication with a page teletypewriter. Also since commercial service is given at speeds of 40 and 60 words per minute, many of the monitoring machines are equipped with two-speed governors and a switch to provide for changing from one speed to the other. These machines are also usually arranged for normal operation from commercial power supply but emergency operation from the 130-volt telegraph battery.

For making test measurements over circuits a special orientation scale is provided together with a small crank extending through the cover for quickly shifting the orientation setting to any desired point. With the machine carefully adjusted to be practically free of harmful distorting effects on the signals, it may then be used for measuring distortions in received teletypewriter signals, the scale being arranged to read the total distortion directly in percent of a pulse length.

Tape Storage Transmission

A heavy volume of traffic may be transmitted rapidly and conveniently by the use of perforated tape. In this method a machine known as a perforator and having a keyboard like that of the teletypewriter is used for punching the code signals for the message in a strip of paper tape. This may be done with simultaneous typing of the message on the teletypewriter in which case the speed of perforating is limited to the speed for which the teletypewriter is set. If a typed record is not made simultaneously with the perforating, punched tape may be prepared at practically any speed within the capabilities of the operator. This punched tape may then be fed through a device known as a tape transmitter which automatically transmits the message signals from the tape at the maximum speed for which the teletypewriters connected to the circuit are set, which is usually 60 words per minute.

The method of transmitting from perforated tape has the distinct advantage of using the line at maximum efficiency at all times as compared with direct keyboard sending where pauses in operating the keys

and interruptions to the operator result in direct losses of circuit time and effectively slower transmission.

Another important advantage of the perforated tape method is that errors may be corrected in the tape before transmission with the result that only errorless copy is transmitted on the circuit. This is done in the following manner and is illustrated in the section of perforated tape shown below. If the operator in attempting to write the word THE should strike the keys T and J (in error), realizing her error she back spaces the tape one division, strikes the "letters" key and then the correct keys H and E. The transmission of the "letters" signal will cause no operation in the recording teletypewriters since they are already in the "letters" case, and the word will be recorded correctly as though no error had been made. Similarly, entire words or groups of characters may be erased from the tape if desired.

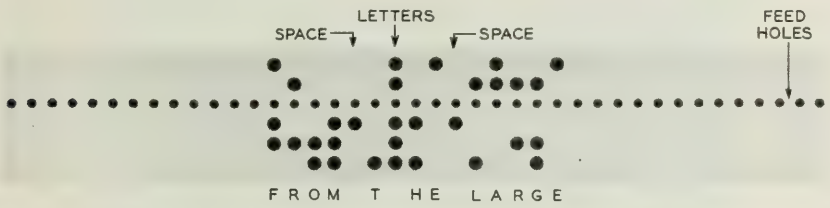


Fig. 9—Sample of strip of perforated tape.

For TWX service a further advantage of the perforated tape method is that the entire message may be punched in tape and checked by printing, if desired, before a call is placed and a connection established. Then, when the connection is established, the message can be automatically transmitted at maximum speed requiring a minimum time for the connection and giving a minimum charge for the call.

It is true, of course, that in this method there is some delay between perforation and transmission. For this reason short to and fro messages, as required in setting up a connection, may be better handled by direct keyboard. To facilitate such working, the perforator keyboard is normally arranged so that by throwing a switch this same keyboard may be used for direct keyboard sending without perforating. This switch also has an intermediate position in which the keyboard is connected for simultaneous direct sending and perforating. This provides for meeting the needs of certain TWX subscribers who wish to simultaneously type and punch the message so that the typed copy may be checked as it is perforated.

The complete page printing set arranged for tape transmission is

known as the No. 19 teletypewriter set and is illustrated in Fig. 10. It employs a No. 15 teletypewriter as the page printing unit.

Fig. 10—No. 19 teletypewriter set.

Automatic Retransmission Using Reperforators

At times it is desirable to retransmit messages received from one circuit to some other machine or machines on a separate circuit. A unit known as a "reperforator" is often used to facilitate such retransmission. The reperforator now standard for the Bell System is a start-stop receiving device using the 5-unit permutation code. It is somewhat similar to the receiving-only tape teletypewriter except that the record produced consists of code perforations in a tape rather

than typing on a tape. This perforated tape is the same as tape produced by a keyboard perforator as previously described, and may be used in an automatic transmitter for retransmitting the message on a separate circuit. The reperforator is usually associated with a receiving teletypewriter on a circuit and may be cut in or out manually or automatically from signals transmitted along with the message signals, so that it will automatically reproduce a code tape for use in automatically retransmitting such messages as desired on some new connection.

Conclusion

The fundamentals of teletypewriters, as described above, now seem to be fairly well established. The future should bring simpler and cheaper machines, especially where the more difficult requirements do not have to be met, and probably additional attachments and auxiliary features to extend the applications and convenience of operation.

The Dielectric Properties of Insulating Materials

By E. J. MURPHY and S. O. MORGAN

This article discusses the variation of dielectric constant and dielectric loss in the radio and power frequency range with the object of giving a simple picture of the type of mechanism which is able to produce anomalous dispersion in this range of frequencies. Some of the general characteristics of anomalous dispersion can be demonstrated as well on a simple and arbitrary model of the structure of dielectrics as on the more complex ones which correspond more closely to the actual structure of dielectrics. Such a derivation is given here in order to indicate the significance of the different factors which occur in the formulæ which have been proposed to account for the variation of dielectric constant and dielectric loss with frequency. This enables a distinction to be made conveniently between the general characteristics which are shared by several types of dielectric polarization and the special characteristics which are peculiar to a restricted class of polarizations or to a particular kind of polarization.

II. Dielectric Polarizability and Anomalous Dispersion

IN a previous paper [1] the general features of the dependence of dielectric constant on frequency were indicated schematically for the entire range extending from the frequencies used in power transmission to those of ultra-violet light. In the range of frequencies below the infra-red (that is, in the electrical range of frequencies) anomalous dispersion is the rule, normal dispersion not having been observed as yet, except for piezo-electric materials, whereas at high optical frequencies normal dispersion is the predominant feature. In the intermediate infra-red region it is not surprising to find a behavior which shows anomalous and normal dispersion in more nearly equal degrees of prominence.

It will be recalled that anomalous dispersion is the type of frequency-variation in which the dielectric constant decreases with increasing frequency, while normal dispersion is the reverse of this, the dielectric constant or refractive index increasing as the frequency increases. The use of the term *anomalous dispersion* to describe the dependence of dielectric constant on frequency in the radio and power frequency range is now widespread, and seems quite appropriate, for it brings out the point that the variation of dielectric constant with frequency in

[1] Murphy and Morgan, *B. S. T. J.*, *16*, 493 (1937).

the radio and power range is in certain respects the same type of phenomenon as optical anomalous dispersion.

Anomalous dispersion plays a very important part in the behavior of dielectrics in the electrical range of frequencies. It is seldom possible to interpret a set of measurements of dielectric constant or other dielectric properties without encountering some manifestation of anomalous dispersion or of the other characteristic types of behavior which follow as corollaries of it.

The two catagories, polarizability and dispersion, include a great deal of the dielectric behavior of insulating materials. This paper will deal primarily with anomalous dispersion, but the theory of anomalous dispersion is not entirely separable from that of the polarizations of which it is an attribute, so it will be necessary to discuss at least briefly the nature of dielectric polarization.

The Relation between Polarizability and Dielectric Constant

For our purposes a dielectric may be thought of as an assemblage of *bound charges*, where this term is intended to include the electrons and positive cores in atoms and molecules, the ions held at lattice points in ionic crystals and, in general, any assemblage of charged particles which are so bound together that they are not able to drift from one electrode to the other under the action of an applied electric field of uniform intensity. Actual dielectrics, of course, also contain some conduction electrons or ions which are free to drift through the material and discharge at the electrodes, producing a direct current conductivity. This conductivity is small at ordinary temperatures in materials classified as dielectrics.

The positions of these charged particles may be considered to be determined by an equilibrium of forces. When an electric field is applied this equilibrium is disturbed and the bound charges are displaced to new positions of equilibrium; then when the applied field is removed they revert to their initial positions. In the equilibrium positions which the charges occupy when a constant electric field has been impressed on the dielectric they have a larger potential energy than in their initial positions. Moreover, they do not revert instantly to their initial positions, and when the retardation is due to friction some of the potential energy of the bound charges is dissipated as heat in the dielectric.

When an alternating voltage is applied to the dielectric, we may think of the bound charges as moving back and forth with certain amplitudes, a different amplitude for each different type of bound charge. When the applied electric field is of unit intensity, the sum of the product of

amplitude and charge extended over all of the *bound* charges in a unit volume of the material determines the *dielectric constant* of the material. The energy dissipated as heat by the motions of these bound charges in the applied electric field represents the *dielectric loss* per second, a quantity which is proportional to the a.-c. conductivity after the d.-c. conductivity has been subtracted from it. The imaginary part of the complex dielectric constant is proportional to the dielectric loss per cycle.

While the physical meaning of the dielectric constant and dielectric loss can be conveniently described, as above, in terms of the amplitudes and energy relationships of bound charges in their motions in an applied electric field, a more useful basis for the discussion is that provided by the concept of polarizability. In the present application the polarizability is equivalent to the product of charge and amplitude, but it has the advantage of being a quantity which is defined and discussed in the general theory of electricity as well as in that of dielectrics. The dielectric constant is then found to be related closely to the polarizabilities of the assemblages of charged particles which the dielectric contains.

The polarization of an assemblage of charges is a quantity defined in electrostatic theory as the vector sum

$$\mathbf{p} = \sum e_i \mathbf{s}_i, \tag{1}$$

where $\mathbf{s}_i$ is the distance of the i^{th} charge, e_i, from a point chosen as origin, and the summation is extended over all of the charges in the assemblage, for which e_i is a typical charge. (If the assemblage has no net charge ($\sum e_i = 0$), the origin may be arbitrarily located without affecting the value of $\mathbf{p}$.)

The polarization is a vector quantity. It can be written as the product of a scalar quantity p, which represents the magnitude or electric moment of the polarization and a unit vector $\mathbf{p}_1$ which gives the direction of the polarization; thus $\mathbf{p} = p\mathbf{p}_1$. As it will not be necessary to distinguish between the properties of isotropic and anisotropic materials in this article the direction of the polarization need not be emphasized. The notation will therefore be simplified, in general, by using the magnitude or scalar part of such vector quantities as the polarization, the electric field intensity and the displacement of charged particles.

To illustrate the application of equation (1) let us consider a very simple configuration consisting of two charges $+ e$ and $- e$ (see Fig. 1). The vector polarization of this configuration is $\mathbf{p} = e(\mathbf{s}_1 - \mathbf{s}_2) = p\mathbf{p}_1$, where p is the magnitude or electric moment of the polarization and

$\mathbf{p}_1$ is a unit vector in the direction of the vector $(\mathbf{s}_1 - \mathbf{s}_2)$. If now one of these charges is an electron ($e = 4.77 \times 10^{-10}$ e.s.u.) and the other a unit positive charge and they are separated by a distance of the order of magnitude of atomic distances (10^{-8} cm.), p has the value 4.77×10^{-18} e.s.u., or 4.77 Debye units. The permanent electric moments of molecules seldom exceed a few Debye units.

Let us now apply the definition contained in equation (1) to a dielectric material. In the first place it indicates that if we know the effective positions of the electrons and other charged particles which

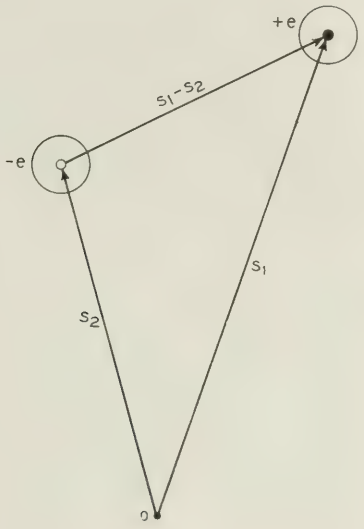


Fig. 1—The calculation of the polarization vector by the general method for a very simple configuration.

contribute to the structure of the material we can always, in principle, calculate the polarization of the body as a whole or any part of it. Actually the calculation of the polarization of a body as a whole or that of unit volume in it is in general a complicated matter involving statistical considerations, but there are special cases in which the result is rather obvious. For example, in a gas or liquid if all orientations of the molecules are equally probable in the absence of an applied field, the value obtained by taking the time-average of the summation indicated by (1) is zero. Equation (1) would also give the value zero when applied to all of the ions in a c.c. of a solution because any arbitrarily chosen small volume in the liquid would be as likely to contain a positive ion as a negative ion.

In some crystalline materials equation (1) gives the value zero because there is a suitable symmetry in the configuration of charged particles in the unit cell; for other solids equation (1) gives a finite value for the unit cell, but zero when applied to a volume of the material large enough to contain a great many crystallites with random orientations; however, there are some macroscopic crystals which have permanent polarizations. A solid material consisting of polar crystallites with random orientations is analogous, as far as equation (1) is concerned, to a liquid or gas containing polar molecules having random orientations; the polarization of the material as a whole is zero in either case.

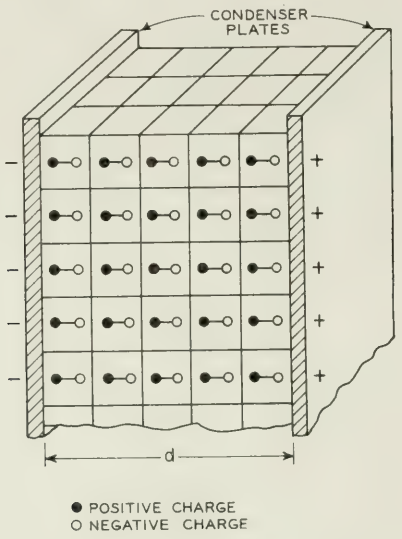


Fig. 2—A dielectric in a condenser. The circles joined by a bar represent "bound charges" of various kinds, including atoms and molecules.

Let us now consider a dielectric of any kind occupying the space between two plane, parallel condenser plates of great enough area and small enough separation that the electric field between the plates when they are charged may be considered to be directed normally to them (cf. Fig. 2). Consider the space between the plates of the condenser to be divided into small cubes of the same size, the purpose of this imaginary division of the dielectric being merely to obtain a representative specimen of the dielectric material. If the cube size is too small the instantaneous value of p obtained by applying equation (1) to all of the particles in a cube will vary appreciably from one cube to another;

but we can then increase the size of the cubes until p is the same for each cube to a close enough approximation. The polarization in each cube is then representative of that of the dielectric as a whole,[2] and by dividing $\sum e_i s_i$ for a typical cube by the volume of the cube we obtain the *polarization per unit volume*, which for the present will be designated as P. This quantity is a statistical mean value involving a summation over a large number of particles; its value depends not only on the structure of the material but upon the effect of thermal motions on the mean positions and orientations of the molecules or other elementary particles in the material. One of the most interesting points in dielectric theory is the consideration—pointed out by Debye and at the basis of his theory of polar molecules—that for some types of structure the mean positions of the particles from which P is calculated are unaffected by changes in the amplitude of thermal motions while for another type of structure (consisting of polar molecules free to assume many or at least several orientations) an increase of temperature decreases P, because the randomness of the orientations of the polar molecules is increased.

For many materials P is zero when no electric field is applied, and assumes a finite value only when an electric field is applied, though as has been indicated, some crystalline materials have a finite value of P even in the absence of an applied electric field. In either case, however, the application of an electric field causes the bound charges within the dielectric to be shifted in general to new equilibrium positions, corresponding to the slight change in the system of forces acting upon them, and if the material did not have a polarization before the application of the field, it assumes one; if it did, it assumes a different value of P. The value of P when an electric field E is applied will be designated as P_E, and that when no field is applied by P_0. Then $P_E - P_0$ is the polarization per unit volume induced by an applied field E. As the dielectric constant of a material depends upon the magnitude of the polarization induced in it by an applied field, and we are concerned here with dielectric constants, it will be desirable to simplify the notation by setting $P_E - P_0 = P$. This gives P a slightly different meaning than it had in the earlier part of the discussion, where it represented the total polarization per unit volume whatever its origin.

[2] A detailed consideration of the method of dividing a dielectric up into elementary volumes in order to compute the mean polarization encounters complications which need not be discussed here. A critical analysis of the method of computing the volume density of polarization of a dielectric is given by Mason and Weaver, "The Electromagnetic Field," Chicago (1929); Chapter III.

The relation between the applied electric field, E, and the polarization induced by it per unit volume is given by

$$P = \frac{\epsilon - 1}{4\pi} E \tag{2}$$

for isotropic materials. The constant $(\epsilon - 1)/4\pi$ is the *susceptibility* of the dielectric in e.s.u., and ϵ is the dielectric constant, which is defined as C/C_0, where C is the capacitance of the measuring condenser while it contains the dielectric and C_0 is its capacitance when empty.

For some purposes there are advantages in considering the actual polarization, which is produced by a discontinuous distribution of charged particles, to be replaced by a vector point function which gives equivalent external effects. Then a vector **P** may be considered to be associated with every point in the space occupied by the dielectric and the dielectric may be considered to have a continuous volume density of polarization,[3] **P**. In non-isotropic bodies the polarization vector **P** induced by an applied field **E** is not always in the same direction as **E**, but is assumed to be a linear vector function [4] of **E** (involving, in the general case, six independent constants), where both **E** and **P** are vector point functions.

In deriving the relationship between the dielectric constant and the molecular structure of a material it cannot be assumed in general that the local field which is impressed upon the elementary particles in the dielectric is simply the field E which can be computed by dividing the applied voltage V by the distance between the plates of the condenser, the intensity of the field being assumed to be uniform. For there is an interaction between the molecules of the dielectric such that each molecule exerts a force on every other molecule. In the absence of an applied electric field these forces combine with other influences to create a distribution for which the polarization per unit volume has the value P_0 (frequently zero, as has been mentioned). Then when a field is applied each element of volume in the dielectric is put into a polarized condition and in general the forces which it exerts upon the particles in other volume elements changes, because the charges in each volume element have been displaced to new positions. Consequently, the value assumed by P in a given cube of Fig. 2 will depend not only upon the direct action of the charges on the plates of the condenser—which determines the strength of the field E—but also

[3] Cf. Mason and Weaver, loc. cit. Chap. III.

[4] Cf. P. Debye, "Polar Molecules," Chemical Catalogue Co., New York (1929), pp. 32–35.

upon their indirect action through the polarization which they create in other elements of volume.

The contribution which the polarization of the dielectric makes to the force upon a charged particle in it has been calculated by Lorentz to be $(4\pi/3)P$, where P is the polarization per unit volume induced by the applied field. This calculation applies to an array of particles with cubic symmetry and to isotropic materials.[5] The *internal or local field* F is then given by

$$F = E + \frac{4\pi}{3} P. \tag{3}$$

E may be thought of as the force which has its origin in the direct interaction between the charges on the plates of the condenser and the charges in the polarizable complex on which attention has been fixed (such as one of the cubes of Fig. 2), while the term $(4\pi/3)P$ may be regarded as an indirect force coming from the other parts of the dielectric by virtue of their polarized state.

It is assumed in the theory of dielectrics that the structure of materials is such that P is a linear function of F (or a linear vector function in the case of anisotropic materials); then

$$P = kF, \tag{4}$$

where k is the polarizability per unit volume. It can be seen that

$$F = \frac{E}{1 - Ak}, \tag{4a}$$

where $A = 4\pi/3$, and consequently that the relation between the polarizability k and the susceptibility $(\epsilon - 1)/4\pi$ $(\equiv K)$ is

$$K \equiv \frac{\epsilon - 1}{4\pi} = \frac{k}{1 - Ak}, \tag{4b}$$

whenever (3) is a valid expression for the internal field.

The susceptibility can be calculated without presupposing the validity of equation (3) for the internal field, *while the value of k depends upon whether (3) or some other expression gives the strength of the internal field in the dielectric.*

If L is the number of molecules per cubic centimeter, $k/L(\equiv \alpha)$ is the polarizability per molecule. This molecular constant α is called the polarizability of the molecule. By multiplying α by Avogadro's number N, we obtain the polarizability per mole of the dielectric:

[5] H. A. Lorentz, "The Theory of Electrons," p. 138, and Notes 54 and 55.

$N\alpha = Nk/L$. And if m is the mass of a molecule, $Nm = M$, where M is the molecular weight, and $Lm = \rho$, where ρ is the density; so that

$$\frac{N}{L} = \frac{M}{\rho}$$

and the polarizability per mol may be written as Mk/ρ.

From equations (3) and (4) (or 4*b*) it can be shown that the polarizability is related to the dielectric constant by the familiar relation

$$k = \frac{3}{4\pi}\left[\frac{\epsilon - 1}{\epsilon + 2}\right], \tag{5}$$

which however is only valid when (3) is valid—and for some materials (3) is apparently not valid.

For gases the term $(4\pi/3)P$ in (3) is so small as compared with E that F is approximately equal to E and

$$k = K = \frac{\epsilon - 1}{4\pi}. \tag{6}$$

The polarizability and susceptibility are then equal. The physical reason for this is that the ratio of intermolecular space to the space occupied by molecules is much larger in a gas than in a solid or liquid and the direct force exerted by the charges on the condenser plates on a charged particle in the dielectric is then much greater than the indirect force which they exert through the polarization induced in other molecules.

It is customary to call the quantity $(4\pi/3)k$ the volume polarization, and it is often denoted by the letter p. The volume polarization may be thought of as $4\pi/3$ times the polarization induced in the dielectric per unit volume per unit applied field. The convenience of using $(4\pi/3)k$ instead of k comes from the occurrence of the factor $4\pi/3$ in the relation (5) between dielectric constant and polarizability.

On dividing equation (5) by the density we obtain a quantity which is called the mass polarization, as it is $4\pi/3$ times the polarizability per gram:

$$\frac{1}{\rho}\left[\frac{\epsilon - 1}{\epsilon + 2}\right] = \frac{4\pi}{3}\frac{k}{\rho}. \tag{7}$$

And on multiplying (7) by the molecular weight of the material we obtain

$$\frac{M}{\rho}\left[\frac{\epsilon - 1}{\epsilon + 2}\right] = \frac{4\pi}{3}\cdot\frac{M}{\rho}\cdot k = \frac{4\pi}{3}\frac{kN}{L} = \frac{4\pi}{3}N\alpha. \tag{8}$$

The quantity $(4\pi/3)N\alpha$ is the *molar polarization*, $N\alpha$ being the polarizability per mole.[6]

Equation (8), and also (7), expresses the Clausius-Mosotti relation when α is considered to be a constant characteristic of the individual molecule and independent of density. The function of ϵ on the left-hand side of (8) is independent of density whenever α is independent of density.

The following relation, analogous to that of Clausius and Mosotti but expressed in terms of the refractive index n, was derived by Lorentz and by Lorenz:

$$\frac{M}{\rho}\frac{n^2-1}{n^2+2}=\frac{4\pi}{3}N\alpha. \tag{8a}$$

The left-hand member of this equation is called the *molar refraction*. Equations (8) and (8*a*) are equivalent because of the general relation between refractive index and dielectric constant ($n^2 = \epsilon$), but owing to the fact that refractive indices are measured at optical frequencies the molar refraction contains only the electronic part of the total molar polarization of the material. Subtracting the molar refraction from the total molar polarization, is one of the methods of determining the amount of polarization contributed by non-electronic polarizations.

It has been found that the Clausius-Mosotti relation is not equally satisfactory for all kinds of dielectric polarization. It gives good results when applied to electronic and atomic polarizations. For example, in an interesting paper on materials of high dielectric constant, Frank [7] has recently shown that the Clausius-Mosotti-Lorentz-Lorenz relationship aids materially in explaining the behavior of the dielectric constants of crystalline materials of high dielectric constant where the dielectric constant depends upon electronic polarizations. Where the polarizability of a molecule is the sum of the polarizabilities of the atoms of which it is composed it is to be expected that if the relation (5), or (8) or (8*a*) is valid the sum of the atomic polarizations would be equal to the molar polarization. Experimental agreement

[6] The polarizabilities of non-polar molecules and atoms are usually of the order of magnitude of 10^{-24} c.c., and the molar polarizations of such substances, consequently, are of the order of magnitude of a few c.c., since the molar polarization is $(4\pi/3) \times 6.06 \times 10^{23}$ times the polarizability of the individual molecule. The polarizability of a conducting sphere is equal to the cube of its radius. And, as atomic dimensions are of the order of magnitude of 10^{-8} cm., it is evident that the polarizabilities of atoms tend to be of a similar order of magnitude to the polarizabilities which would be expected if they behaved as conducting spheres, though there are large differences in the ratio of polarizability to volume for different atoms. The molar polarizations of polar molecules are in general larger than those of similar non-polar molecules and may be a few hundred c.c. (Cf. P. Debye, "Polar Molecules," pp. 12–19.)

[7] F. C. Frank, *Trans. Faraday Society, 23*, (4), 513 (1937).

with this requirement has been found in optics where the refractive indices of molecules can be calculated approximately from the molar refraction (eq. 8*a*) obtained by adding the atomic refractions.

This additive property of electronic polarizations has been employed by Frank [8] to interpret the tendency of crystalline materials having high dielectric constants to be characterized by a high polarizability/volume ratio for the atoms or ions of which they are composed. This condition would tend to allow the largest number of highly polarizable particles to be concentrated in a given space, giving, on the additivity rule, a high molar polarization and a high dielectric constant.

On the other hand Wyman [9] has pointed out that the Clausius-Mosotti relation is not satisfactory when applied to *highly polar liquids*, such as water, and has found that for these substances it appears to be more satisfactory to consider that the polarization is related to the dielectric constant by the empirical relation

$$\frac{\epsilon + 1}{8.5} = \frac{4\pi}{3} k. \tag{9}$$

The calculation of the internal field by Lorentz, which provides the theoretical basis for equation (8), was made before the theory of polar molecules had been developed, but equation (8) has since been applied tentatively to polar molecules.[10] The problem of obtaining an improved relationship between polarizability and dielectric constant for materials having molecules with permanent electric moments has been studied in recent years by several investigators.[11] The calculation of the internal field usually involves the assumption that the effect of the molecules included in a small sphere surrounding the central molecule on which the force is being calculated is negligible on the average because of the random motions due to thermal agitation. On the supposition that such an assumption is not justified in a polar material because of the interactions of adjacent polar molecules, Onsager [12] has obtained a relation between polarizability and dielectric constant which for high dielectric constants is nearly the same as Wyman's empirical relation, equation (9). A comprehensive study of the effects of interaction between the dipoles of polar molecules has

[8] Loc. cit.
[9] Cf. Wyman, *Jour. Amer. Chem. Soc.*, *56*, 539 (1934); *58*, 1482 (1936).
[10] Cf. Debye, loc. cit., p. 13.
[11] Cf. Onsager, *Jour. Amer. Chem. Soc.*, *58*, 1486 (1936); Van Arkel and Snoek, *Trans. Faraday Soc.*, *30*, 707 (1934); Wyman, *Jour. Amer. Chem. Soc.*, *58*, 1482 (1936); Van Vleck, *Jour. Chem. Physics*, *5*, 320 (1937) and *5*, 556 (1937).
[12] Loc. cit.

been made by Van Vleck by the methods of statistical mechanics. He obtains an expression which agrees to a second approximation with that obtained by Onsager. Thus it seems that for highly polar liquids the relations between polarization and dielectric constant developed by Onsager, Wyman and Van Vleck may be more satisfactory than the Clausius-Mosotti relationship, though for many other materials the Clausius-Mosotti relationship is apparently valid or approximately valid.

In deriving expressions for the dependence of dielectric constant on frequency later in this article the formulæ obtained will naturally depend upon which of the equations, (5), (6) or (9), is taken as the relationship between polarizability and dielectric constant. The alternative expressions will be listed.

Derivation of a Dispersion Formula

The above-described relations between polarization and dielectric constant provide the means of obtaining expressions for the variation of dielectric constant with frequency when we have determined the dependence of polarizability on frequency. As our object is to exhibit the general features of anomalous dispersion shared by several particular types of polarization, it will be sufficient to derive dispersion formulæ containing constants the values of which are not specified, but which have a sufficiently obvious physical significance. The derivation given will parallel that of Lorentz in deriving a formula for optical dispersion,[13] and in fact is simply a special case of it in which certain terms are considered to be negligible by comparison with others.

An analogous procedure was used in one of the earliest attempts to explain anomalous dispersion in the electrical frequency range, the theory proposed by Drude [14] in 1898. This theory was based upon the hypothesis that anomalous dispersion in the electrical frequency range depends upon a mechanism similar to that to which optical dispersion was attributed, the difference being that the particles which produce anomalous dispersion in the electrical frequency range are so large that some of the terms in the optical dispersion formula can be neglected. The formulæ which Drude derived for electrical anomalous dispersion yield the same form of variation of dielectric constant with frequency as do the generally accepted theories of the present time, such as the Debye theory; the differences lie in the expressions given for the con-

[13] H. A. Lorentz, "The Theory of Electrons," Chapter IV. See also Korff and Breit, *Reviews of Modern Physics*, *4*, 471 (1932), where a review of the classical theory of optical dispersion is given.

[14] P. Drude, *Ann. d. Physik*, *64*, 131 (1898), "Zur Theorie der anomalien elektrischen Dispersion."

stants in the formulæ in terms of properties of the material. Another adaptation of optical dispersion theory to the explanation of dispersion in the electrical frequency range was proposed by Décombé [15] in 1912. He employed the Lorentz electron theory for the dispersion of light as a basis for the consideration that if the environment of some of the electrons in dielectrics is suitable their motions in an applied field could produce anomalous dispersion and dielectric loss in the electric frequency range. A similar simple and arbitrary assumption regarding the structure of dielectrics will also be employed here. However, it is not proposed as a theory of dielectric behavior but merely employed as a comparatively simple means of deriving and discussing relationships which can be demonstrated as well on a simple and arbitrary model as on the more complex ones which correspond more closely to the actual structure of dielectrics. The relation of the constants in the dispersion formulæ which will be derived here to the actual structure of dielectrics will only be indicated in a general qualitative way for the purpose of illustrating the physical nature of the processes involved; no attempt will be made to provide expressions for the dispersion constants in terms of other observable properties of the material.

In Fig. 2, let the applied potential be V, where V may vary in general in any way with the time, though in the present discussion it will be considered to vary sinusoidally with the time; the impressed field strength is then given by $E = V/d$. As in the more general discussion which preceded this, it will be assumed that the imaginary cells pictured in Fig. 2 contain large numbers of polarizable complexes consisting of positive and negative charges in equal numbers held in position by constitutive forces—the origin of which need not be specified for our present purposes—such that if they are displaced a distance s from their initial positions they will experience a force fs, where f is a constant, tending to restore them to their initial positions; and that while these charges are in motion as a result of the action of the impressed field they experience a frictional force $r\dot{s}$, where r is a constant and $\dot{s}$ is the velocity in the direction of the impressed field; and, finally that their motion is also retarded by an inertia reaction $m\ddot{s}$, proportional to the mass m and the acceleration $\ddot{s}$ of the particles.

The equation of motion for any typical charge e in a polarizable complex having the above-described specifications is

$$m\ddot{s} + r\dot{s} + fs = eF, \tag{10}$$

where F is given by equation (3) in materials to which the Lorentz calculation of the internal field applies, by $F \leqq E$ in the case of gases

[15] L. Décombé, *Journal de Physique*, (5), *3*, 315 (1912).

and by other expressions—which in some cases may approximate either to $F = E$ or to $F = E + (4\pi/3)P$—for still other materials. The quantities F and s are vectors, but for isotropic materials s is in the same direction as F.

If, following the method employed by Lorentz, we write an equation of the form (10) for each charged particle in a physically small volume δ (such as the cubes of Fig. 2), multiply each equation by e, add the equations for all of the particles in δ, and divide by the volume δ, we obtain

$$m\ddot{P} + r\dot{P} + fP = ne^2F, \tag{11}$$

where $P \equiv (1/\delta)\sum es$ and n is the number of charged particles characterized by the constants m, r and f per unit volume. The volume δ may be considered to be that of one of the cubes in Fig. 2. As indicated earlier it should contain a sufficient number of molecules to give a good mean value for P, the polarization per unit volume, but at the same time it should be small enough not to mask significant spatial variations in P.

When the impressed field E is varying sinusoidally with the time at the frequency $\omega/2\pi$, the local or internal field F tending to displace each charged particle in the dielectric will also vary sinusoidally with the time, though in general out of phase with E, if F is given by equation (3), and can be considered to be given by the real part of $F_0e^{i\omega t}$. Under these conditions

$$P = kF_0e^{i\omega t}$$

is a solution of equation (10) for the steady state provided that

$$k = \frac{ne^2}{(ir\omega - m\omega^2 + f)}. \tag{12}$$

k is the polarizability per unit volume and is a complex quantity, since the term $ir\omega$ in the denominator is an imaginary ($i = \sqrt{-1}$).

Equations (10), (11) and (12) apply to a dielectric having a single type of polarization characterized by the constants f, r, m, n and e. But in general an applied field induces several types of polarization simultaneously in a dielectric, and if we assume that it induces w types which are independent of each other, the total polarization per unit volume is given by

$$P = k_1F + k_2F + \cdots k_wF. \tag{13}$$

The total polarizability is then the sum of the individual polarizabilities, or

$$k = \sum_{j=1}^{w} k_j. \tag{14}$$

In this discussion it will be sufficient to consider that the different types of polarization designated by $k_1, k_2 \cdots k_w$ differ from one another only in having different sets of values for the constants of equation (12), designated by the subscripts $1,2,3 \cdots w$; for example, the character of the polarizability k_1 is specified by the set of constants m_1, r_1, f_1 and n_1.

In the first place it is evident that when the frequency of alternation of the voltage applied to the dielectric lies in the radio and power range it is possible to select any number of sets of values of m, r, f which will make the terms $m\omega^2$ and $r\omega$ negligible in comparison with f in the denominator of (12). Let m_1, r_1, f_1 be an example of such a set of constants and let there be n_1 particles per unit volume to which these constants apply. Then for this type of polarization equation (12) reduces to

$$k_1 = \frac{n_1 e^2}{f_1}. \tag{15}$$

This type of polarization is independent of frequency and will be referred to as an instantaneous polarization or an optical polarization. The main representatives of the instantaneous or optical polarizations are the electronic and atomic polarizations, which experience dispersion in the visible and infra-red but which are independent of frequency in the electrical range, and the contribution of this polarizability to the dielectric constant is therefore frequently calculated from refractive index measurements.

A second type of polarization results if we assume that the dielectric we are considering contains a class of particles for which $m\omega^2$ in equation (12) is negligible by comparison with $r\omega$ and with f, but in which $r\omega$ is of the same order of magnitude as f in the electrical range of frequencies. Let m_2, r_2, f_2 be a typical member of this class, the number of such particles per unit volume of the dielectric being n_2. Then for this class of particles equation (12) becomes

$$k_2 = \frac{n_2 e^2}{(ir_2\omega + f_2)}. \tag{16}$$

This expression represents the type of variation with frequency to which the name *anomalous dispersion* is given, and in the preceding paper the type of polarization which produces it was called an absorptive polarization.

It can readily be seen also that neglecting the $m\ddot{s}$ term in (10) or the $m\ddot{P}$ term in (11) leads to the same expression for k, i.e., equation (16), as does neglecting the $m\omega^2$ term in the denominator of (12). So for any member, (r_2, f_2, n_2), of the class of particles which produces

anomalous dispersion, equation (10) reduces to

$$r_2\dot{s} + f_2 s = eF \tag{17}$$

and equation (11) becomes

$$r_2\dot{P} + f_2 P = n_2 e^2 F. \tag{18}$$

Décombé's theory, which has been mentioned earlier, was based upon an equation equivalent in most respects to (18), while Drude's expressions for dispersion were obtained by a method equivalent to neglecting $m\omega^2$ in (12).

Each term in equations (17) and (18) has an evident dynamical significance. Consequently, a physical picture of the essential nature of the anomalous dispersion process is given by equations (17) and (18) even though the values of constants r_2, f_2, n_2 and e are not specified in terms of independently measurable properties of the dielectric. Thus the term $f_2 s$ represents a restoring force tending to return the particles displaced by the impressed field to their initial positions, the constant f_2 acting as a stiffness coefficient; the term $r_2\dot{s}$ acts as a frictional force, r being a measure of the friction experienced by, for example, a moving ion or a rotating polar molecule; and, finally, eF is the driving force tending to displace a particle of charge e. Evidently conditions which are sufficient to produce anomalous dispersion exist whenever the motion of charged particles in an applied field is sufficiently specified by considering the effects of a restoring force proportional to the displacement of the typical particle and of a frictional force proportional to the velocity of the particle in the direction of applied field, as in equation (17). Or, putting it in more general terms, we may say that anomalous dispersion occurs whenever the relation between the polarization per unit volume and the force due to the internal electric field is given by an equation which can be reduced to (18). However, the possibility that anomalous dispersion may also occur under conditions which cannot be described by equation (18) is not excluded by the considerations given here.

A third type of polarization which can be obtained by selecting suitable sets of values for the constants of equation (12) is that in which none of the terms in the denominator of (12) can be neglected in the electrical range of frequencies. Let k_3 be the polarizability for this type of polarization which can then be represented by affixing the subscript 3 to the constants m, r, f and n of equation (12). This type of dispersion includes both the normal and the anomalous types but, as has already been indicated, in the radio and power ranges of fre-

quency examples of a dispersion of this kind have not as yet been observed in dielectrics which are not piezo-electric.[16] It follows then that dielectrics behave as though the inertia of the particles which contribute to dielectric polarization is small enough that the inertia reaction $m\omega^2$ can be neglected in the electrical frequency range. This is an empirical result; the possibility of a polarization of the type k_3 occurring in the electrical frequency range is not excluded by the general theory of dispersion. The higher the frequency of an impressed field the greater should be the likelihood of encountering the type of frequency-variation described by k_3 (or equation (12)), because the prominence of the $m\omega^2$ term increases with the square of the frequency.

The preceding discussion shows that we can write equation (14) in the form

$$k = k_i + k_a, \tag{19}$$

where k is the total polarizability, k_i is the sum of the instantaneous polarizabilities and k_a the sum of the absorptive polarizabilities, that is, of the polarizabilities which vary with frequency according to equation (16). If for simplicity we take the case in which the dielectric has only one representative of k_i and one of k_a, we obtain by substituting the values of k_i and k_a given respectively in (15) and (16),

$$k = \frac{n_1 e^2}{f_1} + \frac{n_2 e^2}{(ir_2\omega + f_2)} \tag{20}$$

as an expression for the total polarizability.

Defining τ' by $\tau' \equiv r/f$, and dropping the subscripts in (20) to make the notation simpler, we obtain

$$k = k_i + \frac{ne^2}{f}\left[\frac{1}{1 + i\omega\tau'}\right], \tag{21}$$

which is the total polarizability per unit volume for a dielectric having two types of polarization, the one represented in (21) by the instantaneous polarizability k_i and the other by the absorptive polarizability

[16] Piezo-electric crystals such as quartz and Rochelle salt form exceptions, but for them dielectric polarization is coupled to macroscopic mechanical strains in the material and the mass reactance is due to the flexing or extension of the entire crystal. The dielectric constant of such a crystal as measured in almost any direction, shows an increase with increasing frequency, followed by anomalous dispersion. This is the behavior required by equation (12), or rather by an equation for the dielectric constant derivable from equation (12). This dispersion, however, depends upon the size and shape of the crystal, the nature of the electrodes and the manner of supporting the crystal during the measurements, and the exact interpretation of such measurements is a rather complex procedure. See, for example, W. P. Mason, *Proc. I. R. E.*, *23*, 1252–1263 (1935).

specified by the second term on the right. The quantity τ' is called the *relaxation-time*.

On multiplying the left-hand side of equation (21) by $(4\pi/3)(M/\rho)$ and the right-hand side by $(4\pi/3)(N/L)$ we obtain

$$\frac{4\pi}{3}\frac{Mk}{\rho} = \frac{4\pi}{3}N\left[\frac{k_i}{L} + \frac{ne^2}{fL}\left(\frac{1}{1+i\omega\tau'}\right)\right], \tag{22}$$

which is the molar polarization.

For dielectrics to which the Clausius-Mosotti relation applies, equation (8) shows that

$$\frac{4\pi}{3}\frac{Mk}{\rho} = \frac{M}{\rho}\frac{\epsilon - 1}{\epsilon + 2} \tag{22a}$$

and in fact the expression on the right-hand side of (22*a*) is frequently called the molar polarization. Reference to equation (6) shows, however, that for gases (22*a*) reduces to the simpler relation.

$$\frac{4\pi}{3}\frac{Mk}{\rho} = \frac{M}{\rho}\frac{(\epsilon - 1)}{3}. \tag{22b}$$

And for Wyman's relation between dielectric constant and polarizability, which has been discussed earlier, the molar polarization becomes

$$\frac{4\pi}{3}\frac{Mk}{\rho} = \frac{M}{\rho}\frac{\epsilon + 1}{8.5}. \tag{22c}$$

Equations (22*a*), (22*b*) and (22*c*) are not the only relations between dielectric constant and molar polarization which have been proposed, but they apparently cover moderately well many of the conditions met in practice. For the right-hand member of equation (22) can be substituted whichever of the three expressions (22*a*), (22*b*), (22*c*) seems the most suitable for the type of dielectric under investigation.

If in equation (21) ω is set equal to zero we obtain the zero-frequency (or static) polarizability

$$k_0 = k_i + ne^2/f \tag{23}$$

and if ω is set equal to infinity we obtain

$$k_\infty = k_i. \tag{24}$$

Subtraction gives

$$k_0 - k_\infty = ne^2/f. \tag{25}$$

Substituting (24) and (25) in (21) gives

$$k = k_\infty + \left(\frac{k_0 - k_\infty}{1 + i\omega\tau'}\right). \tag{26}$$

The constants ne^2/f and k_i are not present in (26), being replaced by two special values of the polarizability, the zero-frequency value and the infinite-frequency value. However, it is not the polarizability but the dielectric constant which is directly observed in measurements on dielectrics, so it is desirable to replace k_0 and k_∞ by their equivalents in terms of the dielectric constant. But, as the earlier discussion has indicated, the relation between dielectric constant and polarizability is different for different types of dielectrics; three alternative expressions analogous to (22*a*), (22*b*) and (22*c*) will therefore be derived.

For materials to which equation (22*a*) (or the equivalent and simpler relation (5)) applies

$$k_0 - k_\infty = \frac{3}{4\pi}\left[\frac{\epsilon_0 - 1}{\epsilon_0 + 2} - \frac{\epsilon_\infty - 1}{\epsilon_\infty + 2}\right] = \frac{9(\epsilon_0 - \epsilon_\infty)}{4\pi(\epsilon_0 + 2)(\epsilon_\infty + 2)}, \quad (27)$$

where ϵ_0 is the zero-frequency dielectric constant and ϵ_∞ is the infinite-frequency dielectric constant. Then equation (26) can be replaced by

$$\frac{4\pi}{3}k = \frac{\epsilon_0 - 1}{\epsilon_\infty + 2} + \left[\frac{\epsilon_0 - 1}{\epsilon_0 + 2} - \frac{\epsilon_\infty - 1}{\epsilon_\infty + 2}\right]\frac{1}{1 + i\omega\tau'}. \quad (28)$$

By rationalizing and using the second expression given for $k_0 - k_\infty$ in equation (27) we can write equation (28) in the alternative form

$$\frac{4\pi k}{3} = \frac{\epsilon_\infty - 1}{\epsilon_\infty + 2} + \left[\frac{3(\epsilon_0 - \epsilon_\infty)}{(\epsilon_0 + 2)(\epsilon_\infty + 2)}\right]\cdot\frac{1}{1 + \omega^2\tau'^2} - i\left[\frac{3(\epsilon_0 - \epsilon_\infty)}{(\epsilon_0 + 2)(\epsilon_\infty + 2)}\right]\cdot\frac{\omega\tau'}{1 + \omega^2\tau'^2}. \quad (29)$$

Equation (29) is the complex polarizability per unit volume multiplied by the factor $4\pi/3$ and expressed in terms of observable values of the dielectric constant and the relaxation-time τ'. The relaxation-time can also be expressed in terms of the reciprocal of a special value of the frequency; this permits all of the theoretical constants such as ne^2/f and τ' to be replaced by certain special values of the dielectric constant and a critical value of the frequency.

A simpler expression for the polarizability is obtained in the case of gases, or whenever equation (6) gives the relation between polarizability and dielectric constant. Equation (26) then gives

$$\frac{4\pi k}{3} = \frac{1}{3}\left(\epsilon_\infty - 1 + \frac{\epsilon_0 - \epsilon_\infty}{1 + i\omega\tau'}\right). \quad (30)$$

And for materials to which the relation (cf. equation (9)) proposed by

Wyman applies the procedure followed above yields

$$\frac{4\pi k}{3} = \frac{1}{8.5}\left(\epsilon_\infty + 1 + \frac{\epsilon_0 - \epsilon_\infty}{1 + i\omega\tau'}\right). \tag{31}$$

On multiplying equations (29), (30) and (31) by M/ρ three alternative formulæ for the molar polarization of a dielectric having polarizations of the type specified by equation (21) are obtained; the constants in these formulæ include only special values (ϵ_0 and ϵ_∞) of the dielectric constant and the relaxation-time, all of which can be obtained from dispersion curves.

The quantity $k_0 - k_\infty$ is a constant of the material, which, as equation (26) shows, represents the largest value which the absorptive part of the total polarizability, i.e., the k_a term in (19), can have for a given material; it may be described as the zero-frequency or static value of the absorptive part of the polarizability. Evidence as to the nature of a polarization can be obtained by investigating experimentally the dependence of $(k_0 - k_\infty)/\rho$ on temperature; for example, if the polarization is due to the changes of orientation of polar molecules according to the Debye theory this quantity should increase linearly with the reciprocal of the absolute temperature. It is useful, therefore, to express $(k_0 - k_\infty)/\rho$ in terms of observable values of the dielectric constant so that it may be plotted against temperature. In this connection there is, however, the same complication which has appeared in other places in this discussion regarding the relation between dielectric constant and polarizability. The three relations which have been discussed here yield for $(k_0 - k_\infty)/\rho$ the following expressions:

$$(k_0 - k_\infty)/\rho = \frac{3}{4\pi\rho}\left[\frac{\epsilon_0 - 1}{\epsilon_0 + 2} - \frac{\epsilon_\infty - 1}{\epsilon_\infty + 2}\right] \text{(Clausius-Mosotti)} \tag{32a}$$

$$= \frac{3}{4\pi\rho}\left(\frac{\epsilon_0 - \epsilon_\infty}{3}\right) \qquad \text{(for gases)} \tag{32b}$$

$$= \frac{3}{4\pi\rho}\left(\frac{\epsilon_0 - \epsilon_\infty}{8.5}\right) \qquad \text{(Wyman).} \tag{32c}$$

The Complex Dielectric Constant

As the dielectric constant (ϵ) is the quantity directly measured in experimental investigations it is desirable to determine how it should vary with frequency for the type of dielectric polarization described in equation (21) or (29). Solving equation (5) for ϵ we obtain

$$\epsilon = \frac{1 + 8\frac{\pi}{3}k}{1 - 4\frac{\pi}{3}k}. \tag{33}$$

By substituting the expression for $4(\pi/3)k$ given in equation (28) into (33) we obtain

$$\epsilon = \frac{\dfrac{\epsilon_0}{\epsilon_0 + 2} + i\dfrac{\epsilon_\infty}{\epsilon_\infty + 2}\,\omega\tau'}{\dfrac{1}{\epsilon_0 + 2} + i\dfrac{1}{\epsilon_\infty + 2}\,\omega\tau'} \tag{34}$$

or

$$\epsilon = \frac{\dfrac{\epsilon_0}{\epsilon_0 + 2}\left(1 + i\dfrac{\epsilon_0 + 2}{\epsilon_\infty + 2}\cdot\dfrac{\epsilon_\infty}{\epsilon_0}\,\omega\tau'\right)}{\dfrac{1}{\epsilon_0 + 2}\left(1 + i\dfrac{\epsilon_0 + 2}{\epsilon_\infty + 2}\,\omega\tau'\right)}. \tag{34a}$$

Then, by setting

$$\frac{\epsilon_0 + 2}{\epsilon_\infty + 2}\tau' = \tau, \tag{35}$$

we obtain

$$\frac{\epsilon}{\epsilon_0} = \frac{\left(1 + i\dfrac{\epsilon_\infty}{\epsilon_0}\,\omega\tau\right)}{1 + i\omega\tau} \tag{36}$$

and transforming this into polar form to facilitate division gives

$$\frac{\epsilon}{\epsilon_0} = \frac{\rho_1 e^{i\varphi_1}}{\rho_2 e^{i\varphi_2}} = \frac{\rho_1}{\rho_2}\,e^{i(\varphi_1 - \varphi_2)} = \frac{\rho_1}{\rho_2}\Big[\cos(\varphi_1 - \varphi_2) + i\sin(\varphi_1 - \varphi_2)\Big], \tag{37}$$

where

$$\rho_1 = \left[1 + \left(\frac{\epsilon_\infty}{\epsilon_0}\right)^2\omega^2\tau^2\right]^{\frac{1}{2}}, \qquad \rho_2 = \left[1 + \omega^2\tau^2\right]^{\frac{1}{2}},$$

$$\varphi_1 = \tan^{-1}\frac{\epsilon_\infty}{\epsilon_0}\,\omega\tau \qquad \text{and} \qquad \varphi_2 = \tan^{-1}\omega\tau.$$

Equation (37) then gives

$$\frac{\epsilon}{\epsilon_0} = \frac{1 + \dfrac{\epsilon_\infty}{\epsilon_0}\,\omega^2\tau^2}{1 + \omega^2\tau^2} + i\frac{\left(\dfrac{\epsilon_\infty}{\epsilon_0} - 1\right)\omega\tau}{1 + \omega^2\tau^2} \tag{38}$$

or

$$\frac{\epsilon}{\epsilon_0} = \frac{\epsilon_\infty}{\epsilon_0} + \frac{1 - \dfrac{\epsilon_\infty}{\epsilon_0}}{1 + \omega^2\tau^2} + \frac{i\left(\dfrac{\epsilon_\infty}{\epsilon_0} - 1\right)\omega\tau}{1 + \omega^2\tau^2}, \tag{38a}$$

from which we obtain

$$\epsilon = \epsilon_\infty + \frac{\epsilon_0 - \epsilon_\infty}{1 + \omega^2\tau^2} - \frac{i(\epsilon_0 - \epsilon_\infty)\omega\tau}{1 + \omega^2\tau^2}. \tag{39}$$

Equation (39) is *the complex dielectric constant* expressed in rectangular form for a dielectric having a polarizability (per unit volume) given by (21) and in which the internal field (F) is such that the Clausius-Mosotti relation (equation (8)) applies.

For gases the derivation of the expression for the complex dielectric constant from that for the polarizability is simpler, though the same in principle, as the above. From (22*b*) or (6) we see that $\epsilon = 1 + 4\pi k$; and on substituting (30) for $4\pi k$ and rationalizing we obtain

$$\epsilon = \epsilon_\infty + \frac{\epsilon_0 - \epsilon_\infty}{1 + \omega^2\tau'^2} - i\,\frac{(\epsilon_0 - \epsilon_\infty)\omega\tau'}{1 + \omega^2\tau'^2}. \tag{40}$$

And when the relation between polarizability and dielectric constant is that proposed by Wyman, cf. (9) or (22*c*), we again obtain (40) on substituting (31) for $(4\pi/3)k$ in (9) and rationalizing.

It will be noticed that the difference between (40) and (39) is that τ' appears in the former and τ in the latter, τ being given by (35). This shows that the factor $(\epsilon_0 + 2)/(\epsilon_\infty + 2)$ has its origin in the fact that for the conditions to which (39) applies $F = E + (4\pi/3)P$, while for the conditions to which (40) applies $F = E$, or is a linear function of E. τ is the relaxation-time for the dielectric constant, while τ' is the relaxation-time for the polarizable units in the material; when $F = E$ these two relaxation-times are equal.

For materials of high dielectric constant the factor $(\epsilon_0 + 2)/(\epsilon_\infty + 2)$ produces a considerable difference between τ and τ'; for example, for water or ice τ is about $23\tau'$. In a recent paper, R. H. Cole [17] has shown that when the volumes of certain polar molecules are calculated from τ by means of Debye's expression for the relaxation-time better agreement with the volume estimated from van der Waals' equation is obtained when Onsager's relation between polarization and dielectric constant is used instead of the Clausius-Mosotti relation. In particular, for water the van der Waals coefficient gives 13×10^{-24} c.c. for the volume of the molecule, while τ' gives 0.5×10^{-24} c.c. on the Clausius-Mosotti relation but 12×10^{-24} c.c. on the Onsager relation, and 4×10^{-24} c.c. for a modified Onsager relation. And if Wyman's relation is used, $\tau = 23\tau'$, and the volume should be 23 times that calculated on the basis of the Clausius-Mosotti relationship, or about 11×10^{-24} c.c.

Both equation (39) and equation (40) can be expressed in the form

$$\epsilon = \epsilon' - i\epsilon'', \tag{41}$$

[17] R. H. Cole, *Jour. Chem. Phys.*, *6*, 385 (1938).

where

$$\epsilon' = \epsilon_\infty + \frac{\epsilon_0 - \epsilon_\infty}{1 + \omega^2\tau^2}, \tag{41a}$$

$$\epsilon'' = \frac{(\epsilon_0 - \epsilon_\infty)\omega\tau}{1 + \omega^2\tau^2}, \tag{41b}$$

if τ is considered to be given by

$$\tau = \frac{\epsilon_0 + 2}{\epsilon_\infty + 2}\tau',$$

when the complex dielectric constant is given by (39), that is when the Clausius-Mosotti relation applies, and by

$$\tau = \tau',$$

when the material is a gas, or when Wyman's relation between polarizability and dielectric constant applies.

The real part ϵ' of the complex dielectric constant is usually referred to simply as the dielectric constant, while the imaginary part ϵ'' is frequently called the *loss factor*.[18] There are alternative ways of expressing the same property of the material; for example, the tangent of the loss angle, ϵ''/ϵ', is frequently used instead of ϵ''.

Comparison of Dispersion Formulæ

Comparison of (39) or (41), (41*a*), (41*b*) with equation (69), page 97, of Debye's "Polar Molecules" shows that the equation for the complex dielectric constant derived here is identical with that of the Debye theory (in the present notation τ' corresponds to τ in Debye's book). This means that any characteristics which can be derived from equations (41), (41*a*), (41*b*) without specifying the values of the constants ϵ_∞, ϵ_0 and τ are common to at least two types of polarization, that is, to the polarization due to the effect of an applied field on the orientation of polar molecules according to the Debye theory, and to the polarization described by equation (18) or equation (21).

The difference between the formulæ for dispersion derived here and those of the Debye theory are best seen by comparing the expressions which they yield for the molar polarization. On the Debye Theory:

$$\frac{M}{\rho}\frac{\epsilon - 1}{\epsilon + 2} = \frac{4\pi N}{3}\left[\alpha_0 + \frac{\mu^2}{3kT}\left(\frac{1}{1 + i\omega\tau'}\right)\right].$$

[18] After a suggestion made by E. T. Hoch, *B. S. T. J.*, November (1922), and by H. H. Race, *Jour. A. I. E. E.*, *51*, 354 (1932).

The present derivation gives:

$$\frac{M}{\rho}\frac{\epsilon - 1}{\epsilon + 2} = \frac{4\pi N}{3}\left[\frac{k_i}{L} + \frac{ne^2}{fL}\left(\frac{1}{1 + i\omega\tau'}\right)\right].$$

The constant α_0 has the same significance as k_i/L, only the notation being different; both represent the optical frequency, or "instantaneous," polarizability; μ is the permanent electric moment of the molecule, k Boltzmann's constant and T the absolute temperature. The quantity ne^2/fL which corresponds to $\mu^2/3kT$ in the Debye formula contains three constants n, e, and f whose physical significance is indicated only in a general way (see Appendix). τ' (or in Debye's notation τ) is equal to $4\pi\eta a^3/kT$ in the Debye theory while in the formula derived here $\tau' = r/f$ where r is a frictional coefficient whose physical origin is not specified. Thus though the formula for molar polarization derived here is not directly useful as a means of investigating the molecular (or other) origin of dielectric polarizations, it facilitates distinguishing those aspects of the Debye formula which are peculiar to a polarization depending upon changes in the orientation of polar molecules which are free to assume any (or at least more than one) orientation from the more general aspects shared by other types of polarization, such as the one specified by (18) and (21). Thus, the functions $\mu^2/3kT$ and $4\pi\eta a^3/kT$ are peculiar to the Debye theory, while the function $(1 + i\omega\tau)^{-1}$ also appears in the dispersion formula derived here, as well as in other formulæ to be discussed below.

We have seen that the viscous-elastic type of polarization specified by (18) and (21) produces a complex dielectric constant given by (39), or by the equivalent equations (41), (41*a*), (41*b*), where

$$\tau = \frac{\tau'(\epsilon_0 + 2)}{(\epsilon_\infty + 2)},$$

and that the same formulæ express the complex dielectric constant on the Debye theory of polar molecules when the constants τ' and $\epsilon_0 - \epsilon_\infty$ (or $k_0 - k_\infty$), are given the values derived for them on the Debye theory. Other theories have been proposed to explain the variation of dielectric constant and dielectric loss with frequency, but for the most part these have been derived for composite dielectrics, consisting of two or more layers of different materials, or of small spheres of one material dispersed or embedded in another material. These theories also yield formulæ (41), (41*a*) and (41*b*) for the complex dielectric constant, the expressions for τ' and $\epsilon_0 - \epsilon_\infty$ being, of course,

different from those of the Debye theory.[19] These expressions are included in Table I.

TABLE I

CONSTANTS OF THE FORMULA FOR THE COMPLEX DIELECTRIC CONSTANT EQUATIONS (41), (41*a*) AND (41*b*)

Type of Polarization	$\epsilon_0 - \epsilon_\infty$	τ	$k = f(\epsilon)$
1. Orientation of polar molecules (The Debye theory).[1]	$\frac{4\pi}{3}\cdot\frac{(\epsilon_0+2)(\epsilon_\infty+2)}{3}\cdot\frac{L\mu^2}{3kT}$	$\frac{(\epsilon_0+2)}{(\epsilon_\infty+2)}\cdot\frac{4\pi\eta a^3}{kT}$	Eq. (5)
	$4\pi\cdot\frac{L\mu^2}{3kT}$	$\frac{4\pi\eta a^3}{kT}$	Eq. (6)
	$\frac{34\pi}{3}\cdot\frac{L\mu^2}{3kT}$	$\frac{4\pi\eta a^3}{kT}$	Eq. (9)
2. Displacement of changed particles against elastic restoring forces and viscous frictional forces of undetermined origin, as specified in equation (17). (Modification of Drude theory.)[2]	$\frac{4\pi}{3}\cdot\frac{(\epsilon_0+2)(\epsilon_\infty+2)}{3}\cdot\frac{ne^2}{f}$	$\frac{\epsilon_0+2}{\epsilon_\infty+2}\cdot\frac{\gamma}{f}$	Eq. (5)
	$4\pi\cdot\frac{ne^2}{f}$	$\frac{\gamma}{f}$	Eq. (6)
	$\frac{34\pi}{3}\cdot\frac{ne^2}{f}$	$\frac{\gamma}{f}$	Eq. (9)
3. Interfacial or ionic polarizations:			
(*a*) Two-layer dielectric; layer (ϵ_1, γ_1) being of same thickness as layer (ϵ_2, γ_2), (Wagner).[3]	$\frac{(\epsilon_1\gamma_2-\epsilon_2\gamma_1)^2}{(\epsilon_1+\epsilon_2)(\gamma_1+\gamma_2)^2}$	$\frac{\epsilon_1+\epsilon_2}{\gamma_1+\gamma_2}$	
(*b*) Special case of (*a*); high-resistance blocking layer at electrode/dielectric boundary (Joffé).[4]	C_1/C_∞	RC_1	
(*c*) Suspension of spheres (ϵ_1, γ_1) in a medium (ϵ_2, γ_2), where $p \ll 1$ (Wagner),[5] (Gemant).[7]	$\frac{9p(\epsilon_1\gamma_2-\epsilon_2\gamma_1)^2}{(2\epsilon_2+\epsilon_1)(2\gamma_2+\gamma_1)^2}$	$\frac{2\epsilon_2+\epsilon_1}{2\gamma_2+\gamma_1}$	
(*d*) Special case of (*c*); conducting spheres in an insulating medium, where $\epsilon_1 = \epsilon_2$ and $\gamma_2 \ll \gamma_1$ (Wagner).[5]	$3p\epsilon_1$	$3\epsilon_1/\gamma_1$	
(*e*) Special case of (*d*); conducting shells (Miles and Robertson).[6]	$3p\epsilon_1$	$\frac{3\epsilon_1 b}{2\gamma_2 d}$	

[1] Reference numbers in this table refer to list of references at the end of this paper.

Table I contains a list of expressions which, when substituted for $(\epsilon_0 - \epsilon_\infty)$ and τ in (41*a*) and (41*b*), give several formulæ for the complex dielectric constant. Included in this list are most of the formulæ which have been proposed to explain the simplest type of variation of

[19] Cf. Gemant, "Elektrophysik der Isolierstoffe," Berlin (1930).

dielectric constant and loss factor with frequency which is observed in the different classes of materials to which the various items in the table refer. In some cases the original formulæ, as they appear in the literature, have been expressed in terms which do not show an obvious equivalence to (41), (41*a*) and (41*b*), but by re-expressing them in the form (41) and then determining ϵ_0 and ϵ_∞ by letting $\omega = 0$ and ∞, respectively, the list of expressions given in Table I is obtained. It is interesting that theories based on such widely dissimilar physical mechanisms as rotating polar molecules (Item 1, Table I) and a blocking-layer of high resistance at an electrode/dielectric interface (Item 3(*b*), Table I) should yield an identical form of variation with frequency.

For the first two polarizations listed in Table I, the alternative expressions obtained by assuming three alternative relationships between polarizability and dielectric constant are given. By means of Table II the quantities $(k_0 - k_\infty)$ and τ' can be obtained from $(\epsilon_0 - \epsilon_\infty)$

TABLE II

THE RELATIONSHIP BETWEEN $(k_0 - k_\infty)$ AND $(\epsilon_0 - \epsilon_\infty)$ AND BETWEEN τ' AND τ

	$(k_0 - k_\infty)$	τ'
Clausius-Mosotti Relation	$\frac{3}{4\pi} \cdot \frac{3(\epsilon_0 - \epsilon_\infty)}{(\epsilon_0 + 2)(\epsilon_\infty + 2)}$	$\frac{\epsilon_\infty + 2}{\epsilon_0 + 2}\tau$
Gases	$\frac{1}{4\pi} \cdot (\epsilon_0 - \epsilon_\infty)$	τ
Wyman's Empirical Relation	$\frac{3}{4\pi} \cdot \frac{\epsilon_0 - \epsilon_\infty}{8.5}$	τ

and τ in Table I. The resulting expressions can then be substituted for $(k_0 - k_\infty)$ and τ' in equation (26) yielding expressions for the polarizabilities of the different types of polarization listed in Table I. The molar polarization can then be obtained by multiplying (26) by $(4\pi/3)(M/\rho)$. However, in general it is not likely that any useful purpose is to be served by calculating the molar polarization for interfacial polarizations; a more significant quantity would be the polarization per conducting particle, when the polarization is of the type (3*d*), Table I, and the number of conducting particles per unit volume can be estimated.

We have pointed out that a number of theories which have been proposed for the explanation of the variation of dielectric constant with frequency may be expressed in the forms (41*a*) and (41*b*) when the expressions listed in Table I are substituted for $(\epsilon_0 - \epsilon_\infty)$ and τ, but we have not yet indicated how these formulæ agree with experimental data. For such materials as ice (see Fig. 3, for example) and for

certain alcohols and glycols, the experimental points agree fairly closely with the curves obtained by plotting equations (41*a*) and (41*b*) for a suitable choice of the values of the constants.

But for many other dielectrics, particularly non-homogeneous systems or disperse systems such as those listed under Item 3, Table I, the simple dispersion formulæ (41*a*) and (41*b*) often fall very far short of adequately representing the experimental data. Von Schweidler [20] and Wagner [21] have attempted to explain the form of dispersion curves obtained for such materials by postulating that the polarizations

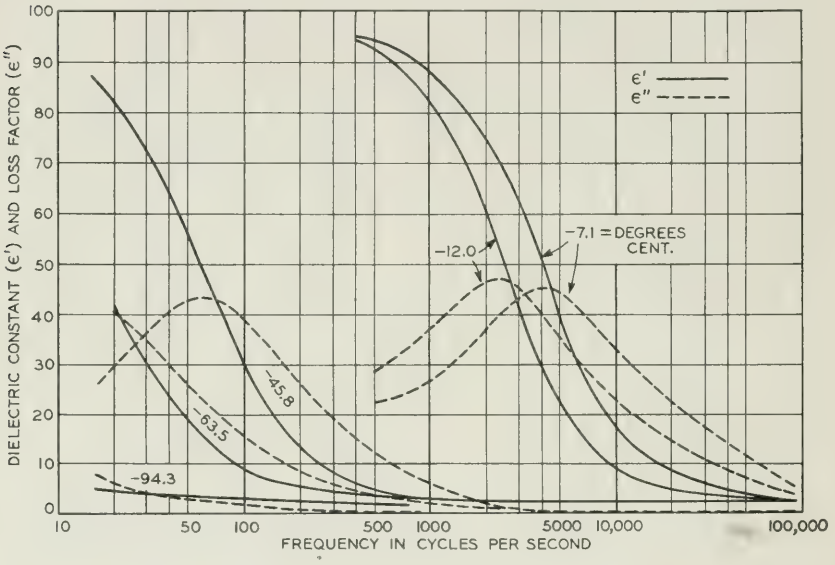


Fig. 3—Experimental dispersion curves for ice.

induced in the dielectric have a wide range of relaxation times at any given temperature, instead of a single relaxation time, as for the polarizations listed in Table I. A further contribution to the theory of the distribution of relaxation times has recently been made by Yager.[22] However, in spite of the existence of many materials which do not show the type of dispersion described by (41*a*) and (41*b*), the value of these formulæ in interpreting experimental data is considerable, particularly as applied to pure materials.

Table I emphasizes the point that mere agreement of experimental data for dielectric constant and dielectric loss with the theoretical

[20] E. v. Schweidler, *Ann. d. Phys.*, (4) (*24*), 711 (1907).
[21] K. W. Wagner, *Archiv f. Elektrotechnik*, *2*, 371 (1914).
[22] W. A. Yager, *Physics*, *7*, 434 (1936).

curves obtained by plotting (41*a*) and (41*b*) for suitably adjusted values of the constants only places the type of mechanism to which the observed dispersion can be attributed within the rather large catagory which includes at least the seven types of mechanism listed in the table. Data showing the dependence of $(k_0 - k_\infty)/\rho$ on temperature allows a further specialization of the processes which could account for the observed behavior; and of course a number of possibilities can be discarded on general grounds of physical improbability. And finally, agreement of the constants calculated from dielectric measurements with the values calculated from independent estimates of the sizes and other characteristics of the molecules or other elementary units which contribute to the polarization provides the most convincing evidence of the nature of the polarization. Such agreement is frequently obtained in the application of the Debye theory to gases and liquids.

The characteristics which can be deduced from equations (41*a*) and (41*b*) without substituting for the constants theoretical expressions, such as those given in Table I, are of considerable value in interpreting electrical measurements upon dielectrics. It may be convenient to describe these as the *general* characteristics of anomalous dispersion, distinguishing them thereby from the *special* characteristics peculiar to particular kinds of dielectric polarization which share the property of producing anomalous dispersion in the radio and power range of frequencies.

Appendix

The following list contains the definitions of the quantities which appear in Table I:

$\epsilon_1, \epsilon_2, \gamma_1, \gamma_2$ are respectively the dielectric constants and conductivities of two materials designated by subscripts 1 and 2, the unit of conductivity being such that $\gamma = 36\pi \times 10^{11}\,\lambda$, where λ is in $(\text{ohm}\cdot\text{cm})^{-1}$.

$\epsilon_0, \epsilon_\infty$ are respectively the dielectric constant at the lower and upper extremities of dispersion curves; they are called the zero-frequency (or static) dielectric constant and the infinite-frequency dielectric constant,

L is the number of molecules per unit volume,

η the viscosity of a liquid containing polar molecules,

k Boltzmann's constant,

T the absolute temperature,

μ the permanent electric moment of a polar molecule,

a the radius of a polar molecule, assumed to be spherical,
b the radius of a colloidal particle,
d the thickness of a conducting skin on the particle of radius b,
r a frictional resistance coefficient of unspecified origin,
f an elastic restoring force coefficient of unspecified origin,
n the number per unit volume of elementary charged particles subject to certain specified conditions,
p the ratio of the volume occupied by the spherical particles in ($3c$, d, e) to the total volume,
C_1 the capacity of the blocking layer of ($3b$), Table I,
R the resistance of the dielectric, exclusive of the blocking layer.

The following list contains the definitions of quantities which appear in other parts of the article.

ω is 2π times the frequency of alternation of the applied field,
V the applied voltage,
E the intensity of the applied field,
P the polarization per unit volume induced by a field E,
F the internal or local field,
ρ the density of the dielectric,
M the molecular weight of the material of which the dielectric is composed,
m the mass of a molecule; in another context, the mass of any charged particle considered in the discussion,
N is Avogadro's number, 6.06×10^{23} molecules per mole,
s the displacement of a charged particle from an equilibrium position by an applied field,
$\dot{s}$ the velocity of the charged particle in the applied field,
$\ddot{s}$ the acceleration of the particle in the applied field.

References Relating to Table I

1. P. Debye, "Polar Molecules," New York (1929).
2. P. Drude, *Ann. d. Physik, 64*, 131 (1898); L. Décombé, *J. d. Physique* (5) *3*, 215 (1912); and the present article.
3. K. W. Wagner, Chap. I of Schering's "Die Isolierstoffe der Elektrotechnik," Springer, Berlin (1934).
4. A. Joffé, "The Physics of Crystals," New York (1928).
5. K. W. Wagner, *Arch. f. Elektrotechnik, 2*, 371 (1914).
6. J. B. Miles and H. P. Robertson, *Phys. Rev., 40*, 583 (1932).
7. A. Gemant, "Die Elektrophysik der Isolierstoffe," Berlin (1930).

General reviews of the theory of dielectric behavior as it concerns dispersion for power and radio frequencies are included in the following places, among others:

1. E. Schrödinger, "Dielektrizität," Graetz, *Handb. d. Elek. u. d. Magn.*, Leipzig (1918), pp. 157–229.

2. E. v. Schweidler, "Die Anomalien der dielektrischen Erscheinungen," *ibid.*, p. 232; *Ann. d. Phys.* (4) *24*, 711 (1907).
3. J. B. Whitehead, "Lectures on Dielectric Theory and Insulation," McGraw-Hill (1927).
4. L. Hartshorn, *Jour. I. E. E.*, *64*, 1152 (1926).
5. P. Debye, "Polar Molecules," Chem. Cat. Co., New York (1929).
6. Schering's "Die Isolierstoffe der Elektrotechnik," Springer, Berlin (1924).
7. A. Gemant, "Die Elektrophysik der Isolierstoffe," Berlin (1930).

Abstracts of Technical Articles from Bell System Sources

Electron Microscope Studies of Thoriated Tungsten.[1] ARTHUR J. AHEARN and JOSEPH A. BECKER. Many past experiments have shown that the thermionic activity of a thoriated tungsten filament is determined by the concentration of thorium on its surface. This concentration is in turn determined by the rate of arrival and rate of evaporation of thorium. Typical published values of these rates are given in Fig. 1. An electron microscope used to obtain electron images of thoriated tungsten ribbons is described. Comparison with photomicrographs shows that the active and inactive patches composing an electron image agree in size, shape and number with the exposed grains of the tungsten. The electron microscope *shows that thorium comes to the surface in "eruptions"* at a relatively small number of randomly located points. From a comparison of photomicrographs showing thoria globules and electron images of thorium eruptions, it is deduced that *all the thorium in a globule comes to the surface when an eruption occurs.* Factors such as a high temperature flash and sudden heating and cooling of the filament affect the frequency of eruptions. Thorium eruptions are the only observed manner in which thorium arrives at the filament surface. They are repeatedly observed in the early stages of thoriation. Eruptions are not observed in the later stages of thoriation where conditions are unfavorable for their observance. Electron images of a Pintsch single crystal filament reveal alternate active and inactive bands parallel to the filament axis. Thorium eruptions occur only on the active bands. With a polycrystalline ribbon the surface migration of thorium from the eruption centers is isotropic; *with a single crystal ribbon there is a strongly preferred direction of migration.* X-ray analysis shows that the surface is a (211) plane and that *the preferred direction of migration agrees with the* (111) *direction in this plane.* During the process of thoriating a filament the relative emissions from different grains change by substantial amounts; in many cases the change is so great that the relative emissions are reversed. Measurements of work function differences between grains gave values ranging up to 0.6 volt.

The Mechanism of Hearing as Revealed through Experiment on the Masking Effect of Thermal Noise.[2] HARVEY FLETCHER. In an electri-

[1] *Phys. Rev.*, September 15, 1938.
[2] *Proc. Nat'l. Acad. Sci.*, July 1938.

cal conductor there is a statistical variation of the electrical potential difference between its two ends which is due to the thermal agitation of the atoms, including the electrons. This electrical noise is amplified by means of a vacuum tube amplifier and then converted into an acoustical noise by means of a telephone receiver held on the ear. When this noise is present it reduces the capability of the ear to hear other sounds. The intensity per cycle of the acoustical noise compared to the intensity of a pure tone which can just be perceived in the presence of a noise was determined experimentally using a group of observers. This relative intensity for a given frequency range was constant throughout a wide variation of intensity. However, its value does vary with the position in the frequency spectrum and it is the amount of this variation which enables one to calculate the relation between the frequency of the tone and its position of maximum stimulation along the basilar membrane. The results of such a calculation are given and shown to be in good agreement with determinations from animal experimentation.

Transcontinental Telephone Lines.[3] J. J. PILLIOD. A fourth transcontinental line has just been created by the completion of four pairs of open wire between Oklahoma City and Whitewater, California. This open-wire line connects at its eastern terminus with the already existing toll cables from the east, and at its western terminus with a toll cable running into Los Angeles.

In a cross-section of the United States just west of Denver, there are now 140 through telephone circuits and about the same number of telegraph circuits carried by four open-wire routes.

The four new pairs which constitute the transcontinental line carry, in addition to the usual voice frequency channels, three channels of carrier. But their design throughout has been such that twelve additional carrier circuits can be superimposed upon the four channels now provided by each wire pair.

The wires of each pair are spaced 8 inches apart with the nearest spacing between pairs being 26″ while crossarms are 36″ apart. New transposition systems have also been used to further reduce crosstalk.

Application of Statistical Methods to Manufacturing Problems.[4] W. A. SHEWHART. The application of statistical methods in mass production makes possible the most efficient use of raw materials and manufacturing processes, effects economies in production, and makes possible the highest economic standards of quality for the manufactured goods used by all of us. The story of the application, however,

[3] *Electrical Engineering*, October 1938.
[4] *Jour. Franklin Institute*, August 1938.

is of much broader interest. The economic control of quality of manufactured goods is perhaps the simplest type of scientific control. Recent studies in this field throw light on such broad questions as: How far can Man go in controlling his physical environment? How does this depend upon the human factor of intelligence and how upon the element of chance?

Observational Significance of Accuracy and Precision.[5] W. A. SHEWHART. Two of the most common terms used in pure and applied science are accuracy and precision. When such terms are used, as in the specification of quality of manufactured products, it is desirable that they have definite and, in so far as possible, experimentally verifiable meanings. It is, therefore, important to determine how far one can go towards attaining this end by applying with rigor the principle that *only that which is observable is significant.* In the application of the concepts of accuracy and precision, it is customarily assumed that the available data constitute a random sample. Hence, the first step in attaining experimentally verifiable meaning of these terms is to choose an operationally verifiable criterion of randomness. One such criterion is the quality control chart. In order to give experimental definiteness to any *measure* of either accuracy or precision derived from a random sample, it is also necessary to specify the way any statement involving the measure may be experimentally verified. To do this it is necessary to make at least four empirical choices as to the details of taking and analyzing the data in the process of verification. Hence, it appears that the meaning of either precision or accuracy is verifiable. Hence, it appears that the meaning of either precision or accuracy is verifiable only in a limited sense subject in any specific case to the choice of empirical criteria of verification.

The Time Lag in Gas-Filled Photoelectric Cells.[6] A. M. SKELLETT. In commercial gas-filled photoelectric cells there is a lag in response which becomes appreciable above frequencies in the neighborhood of 10,000 cycles. If this lag is due to the transit times of the ions across the cell, it should be possible to set up resonance conditions by varying the frequency of modulation of the incident light intensity. This has been accomplished in a cell of special design and the resonance conditions agree with the theory, thereby demonstrating that the transit time of the ions is the simple cause. The paper also discusses the flow of the ions and electrons across the cell and their impacts in relation to the flow of current in the external circuit.

[5] *Jour. Wash. Acad. Sciences,* August 15, 1938 (p. 381).
[6] *Internat'l. Projectionist,* September 1938; *Jour. Applied Physics,* October 1938.

Contributors to this Issue

Charles R. Burrows, B.S. in Electrical Engineering, University of Michigan, 1924; A.M., Columbia University, 1927; E.E., University of Michigan, 1935. Research Assistant, University of Michigan, 1922–23. Western Electric Company, Engineering Department, 1924–25; Bell Telephone Laboratories, Research Department, 1925–. Mr. Burrows has been associated continuously with radio research and is now in charge of a group investigating the propagation of ultra-short waves.

Arthur B. Crawford, B.S. in Electrical Engineering, Ohio State University, 1928. Member of Technical Staff, Bell Telephone Laboratories, 1928–. Mr. Crawford has been engaged chiefly in work relative to radio communication by ultra-short waves.

Carl R. Englund, B.S. in Chemical Engineering, University of South Dakota, 1909; University of Chicago, 1910–12; Professor of Physics and Geology, Western Maryland College, 1912–13; Laboratory Assistant, University of Michigan, 1913–14. Western Electric Company, 1914–25; Bell Telephone Laboratories, 1925–. As Radio Research Engineer Mr. Englund is engaged largely in experimental work in radio communication.

L. A. Meacham, B.S. in Electrical Engineering, University of Washington, 1929. Cambridge University, England, 1929–30. Bell Telephone Laboratories, 1930–. Mr. Meacham's work has been concerned with the generation and distribution of constant reference frequencies.

S. O. Morgan, B.S. in Chemistry, Union College, 1922; M.A., Princeton University, 1925; Ph.D., 1928. Western Electric Company, Engineering Department, 1922–24; Bell Telephone Laboratories, 1927–. Dr. Morgan's work has been on the relation between dielectric properties and chemical composition.

William W. Mumford, B.A., Willamette University, 1930. Bell Telephone Laboratories, 1930–. Mr. Mumford has been engaged in radio receiving work, chiefly on the problem of propagation and measurement in the ultra-short-wave region.

E. J. Murphy, B.S., University of Saskatchewan, Canada, 1918; McGill University, Montreal, 1919–20; Harvard University, 1922–23.

Western Electric Company, Engineering Department, 1923–25; Bell Telephone Laboratories, 1925–. Mr. Murphy's work is largely confined to the study of the electrical properties of dielectrics.

A. C. Norwine, A.B., University of Missouri, 1923; B.S. in Electrical Engineering, 1924; E.E., 1925. Bell Telephone Laboratories, 1925–. Mr. Norwine has been principally engaged in studies of the effects of transmission delay and voice operated devices on toll telephone circuits.

A. J. Rack, B.S. in Electrical Engineering, University of Illinois, 1930; M.A. in Physics, Columbia University, 1935. Bell Telephone Laboratories, 1930–. Starting with radio research, Mr. Rack has more recently been engaged in the analysis of special problems arising in amplifier circuits.

E. F. Watson, M.E., Cornell University, 1914. American Telephone and Telegraph Company, Engineering Department, 1914–19; Department of Development and Research, 1919–34. Bell Telephone Laboratories, 1934–. Mr. Watson has been concerned with the development of various types of telegraph equipment, particularly teletypewriters, telephotograph equipment, telegraph maintenance and testing equipment, grounded telegraph systems and regenerative telegraph repeaters. His present work as Teletypewriter Engineer is along these same lines.

S. B. Wright, M.E. in Electrical Engineering, Cornell University, 1919. Engineering Department and Department of Development and Research, American Telephone and Telegraph Company, 1919–34; Bell Telephone Laboratories, 1934–. Mr. Wright is engaged in transmission development of radio systems.

Index to Volume XVII

A

Almquist, M. L., H. J. Fisher and R. H. Mills, A New Single Channel Carrier Telephone System, page 162.
Amplitude Characteristics of Telephonic Signals, Devices for Controlling, *A. C. Norwine*, page 539.
Amplitude Range Control, *S. B. Wright*, page 520.
Analyzer, An Optical Harmonic, *H. C. Montgomery*, page 406.
Anti-sidetone Subscriber Set, Common Battery, An Explanation of the, *C. O. Gibbon*, page 245.

B

Best, F. H., New Transmission Measuring Systems for Telephone Circuit Maintenance, page 1.
Blanchard, Julian, Hertz, the Discoverer of Electric Waves, page 327.
Bode, H. W., Variable Equalizers, page 229.
Burrows, Chas. R., The Exponential Transmission Line, page 555.

C

Cable Carrier System, Crystal Channel Filters for the, *C. E. Lane*, page 125.
Cable Carrier Telephone System, Crosstalk and Noise Features of, *M. A. Weaver, R. S. Tucker and P. S. Darnell*, page 137.
Cable Carrier Telephone Terminals, *R. W. Chesnut, L. M. Ilgenfritz and A. Kenner*, page 106.
Cable System for Television Transmission, Coaxial, *M. E. Strieby*, page 438.
Cables, Toll, A Carrier Telephone System for, *C. W. Green and E. I. Green*, page 80.
Carr, J. A. and F. V. Haskell, Studies of Telephone Line Wire Spacing Problems, page 195.
Carrier Telephone System for Toll Cables, A, *C. W. Green and E. I. Green*, page 80.
Carrier Telephone System, Cable, Crosstalk and Noise Features of, *M. A. Weaver, R. S. Tucker and P. S. Darnell*, page 137.
Carrier Telephone System, A New Single Channel, *H. J. Fisher, M. L. Almquist and R. H. Mills*, page 162.
Chesnut, R. W., L. M. Ilgenfritz and A. Kenner, Cable Carrier Telephone Terminals, page 106.
Clarke, Beverly L. and A. E. Ruehle, Spectrochemical Analysis in Communication Research, page 381.
Coaxial Cable System for Television Transmission, *M. E. Strieby*, page 438.
Crawford, A. B., C. R. Englund and W. W. Mumford, Ultra-Short-Wave Transmission and Atmospheric Irregularities, page 489.
Crystal Channel Filters for the Cable Carrier System, *C. E. Lane*, page 125.

D

Darnell, P. S., M. A. Weaver and R. S. Tucker, Crosstalk and Noise Features of Cable Carrier Telephone System, page 137.
Darrow, Karl K., Radioactivity—Artificial and Natural, page 292.
Davisson, C. J., The Discovery of Electron Waves, page 475.
Dielectric Properties of Insulating Materials, The, *E. J. Murphy and S. O. Morgan*, page 640.
Diodes, Effect of Space Charge and Transit Time on the Shot Noise in, *A. J. Rack*, page 592.

E

Echo Suppressors, Two, The Occurrence and Effect of Lockout Occasioned by, *Arthur W. Horton, Jr.*, page 258.

Electric Waves, Hertz, the Discoverer of, *Julian Blanchard*, page 327.
Electron Waves, The Discovery of, *C. J. Davisson*, page 475.
Englund, C. R., A. B. Crawford and W. W. Mumford, Ultra-Short-Wave Transmission and Atmospheric Irregularities, page 489.
Equalizers, Variable, *H. W. Bode*, page 229.

F

Fay, C. E., A. L. Samuel and W. Shockley, On the Theory of Space Charge between Parallel Plane Electrodes, page 49.
Feedback Oscillators, Stabilized, *G. H. Stevenson*, page 458.
Filters, Crystal Channel, for the Cable Carrier System, *C. E. Lane*, page 125.
Fisher, H. J., M. L. Almquist and R. H. Mills, A New Single Channel Carrier Telephone System, page 162.

G

Gibbon, C. O., An Explanation of the Common Battery Anti-sidetone Subscriber Set, page 245.
Green, C. W. and E. I., A Carrier Telephone System for Toll Cables, page 80.
Gustafson, W. G., Magnetic Shielding of Transformers at Audio Frequencies, page 416.

H

Haskell, F. V. and J. A. Carr, Studies of Telephone Line Wire Spacing Problems, page 195.
Herriott, W., High Speed Motion Picture Photography, page 393.
Hertz, the Discoverer of Electric Waves, *Julian Blanchard*, page 327.
Horton, Arthur W., Jr., The Occurrence and Effect of Lockout Occasioned by Two Echo Suppressors, page 258.

I

Ilgenfritz, L. M., R. W. Chesnut and A. Kenner, Cable Carrier Telephone Terminals, page 106.
Impedance Concept and its Application to Problems of Reflection, Refraction, Shielding and Power Absorption, The, *S. A. Schelkunoff*, page 17.
Inglis, A. H., Transmission Features of the New Telephone Sets, page 358.
Insulating Materials, The Dielectric Properties of, *E. J. Murphy and S. O. Morgan*, page 640.

J

Jones, W. C., Instruments for the New Telephone Sets, page 338.

K

Kenner, A., R. W. Chesnut and L. M. Ilgenfritz, Cable Carrier Telephone Terminals, page 106.

L

Lane, C. E., Crystal Channel Filters for the Cable Carrier System, page 125.

M

Magnetic Shielding of Transformers at Audio Frequencies, *W. G. Gustafson*, page 416.
Maintenance, Telephone Circuit, New Transmission Measuring Systems for, *F. H. Best*, page 1.
Meacham, L. A., The Bridge Stabilized Oscillator, page 574.
Mills, R. H., H. J. Fisher and M. L. Almquist, A New Single Channel Carrier Telephone System, page 162.
Montgomery, H. C., An Optical Harmonic Analyzer, page 406.
Morgan, S. O. and E. J. Murphy, The Dielectric Properties of Insulating Materials, page 640.
Motion Picture Photography, High Speed, *W. Herriott*, page 393.

Mumford, W. W., C. R. Englund and A. B. Crawford, Ultra-Short-Wave Transmission and Atmospheric Irregularities, page 489.
Murphy, E. J. and S. O. Morgan, The Dielectric Properties of Insulating Materials, page 640.
Murphy, O. J. and A. C. Norwine, Characteristic Time Intervals in Telephonic Conversation, page 281.

N

New Telephone Sets, Instruments for the, *W. C. Jones,* page 338.
New Telephone Sets, Transmission Features of the, *A. H. Inglis,* page 358.
Norwine, A. C., Devices for Controlling Amplitude Characteristics of Telephonic Signals, page 539.
Norwine, A. C. and O. J. Murphy, Characteristic Time Intervals in Telephonic Conversation, page 281.

O

Oscillator, The Bridge Stabilized, *L. A. Meacham,* page 574.
Oscillators, Stabilized Feedback, *G. H. Stevenson,* page 458.

R

Rack, A. J., Effect of Space Charge and Transit Time on the Shot Noise in Diodes, page 592.
Radio: Ultra-Short-Wave Transmission and Atmospheric Irregularities, *C. R. Englund, A. B. Crawford and W. W. Mumford,* page 489.
Radioactivity—Artificial and Natural, *Karl K. Darrow,* page 292.
Research, Communication, Spectrochemical Analysis in, *Beverly L. Clarke and A. E. Ruehle,* page 381.
Ruehle, A. E. and Beverly L. Clarke, Spectrochemical Analysis in Communication Research, page 381.

S

Samuel, A. L., C. E. Fay and W. Shockley, On the Theory of Space Charge between Parallel Plane Electrodes, page 49.
Schelkunoff, S. A., The Impedance Concept and its Application to Problems of Reflection, Refraction, Shielding and Power Absorption, page 17.
Shockley, W., C. E. Fay and A. L. Samuel, On the Theory of Space Charge between Parallel Plane Electrodes, page 49.
Short-Wave, Ultra-, Transmission and Atmospheric Irregularities, *C. R. Englund, A. B. Crawford and W. W. Mumford,* page 489.
Shot Noise in Diodes, Effect of Space Charge and Transit Time on the, *A. J. Rack,* page 592.
Space Charge between Parallel Plane Electrodes, On the Theory of, *C. E. Fay, A. L. Samuel and W. Shockley,* page 49.
Space Charge and Transit Time on the Shot Noise in Diodes, Effect of, *A. J. Rack,* page 592.
Spectrochemical Analysis in Communication Research, *Beverly L. Clarke and A. E. Ruehle,* page 381.
Stevenson, G. H., Stabilized Feedback Oscillators, page 458.
Strieby, M. E., Coaxial Cable System for Television Transmission, page 438.

T

Telephonic Signals, Devices for Controlling Amplitude Characteristics of, *A. C. Norwine,* page 539.
Teletypewriters Used in the Bell System, Fundamentals of, *E. F. Watson,* page 620.
Television Transmission, Coaxial Cable System for, *M. E. Strieby,* page 438.
Time Intervals in Telephonic Conversation, Characteristic, *A. C. Norwine and O. J. Murphy,* page 281.
Transformers at Audio Frequencies, Magnetic Shielding of, *W. G. Gustafson,* page 416.
Transmission Features of the New Telephone Sets, *A. H. Inglis,* page 358.
Transmission Line, The Exponential, *Chas. R. Burrows,* page 555.

Transmission Measuring Systems for Telephone Circuit Maintenance, New, *F. H. Best*, page 1.
Tucker, R. S., M. A. Weaver and P. S. Darnell, Crosstalk and Noise Features of Cable Carrier Telephone System, page 137.

W

Watson, E. F., Fundamentals of Teletypewriters Used in the Bell System, page 620.
Weaver, M. A., R. S. Tucker and P. S. Darnell, Crosstalk and Noise Features of Cable Carrier Telephone System, page 137.
Wire Spacing Problems, Telephone Line, Studies of, *J. A. Carr and F. V. Haskell*, page 195.
Wright, S. B., Amplitude Range Control, page 520.

A. V. EMMOTT & SONS
BOOKBINDERS, Inc.
1101 HAMILTON
HOUSTON 3, TEXAS